Martin A. Leroch
Florian Rupp *Editors*

Power and Responsibility

Interdisciplinary Perspectives for the 21st Century in Honor of Manfred J. Holler

Springer

Power and Responsibility

Martin A. Leroch · Florian Rupp
Editors

Power and Responsibility

Interdisciplinary Perspectives for the 21st Century in Honor of Manfred J. Holler

Springer

Editors
Martin A. Leroch
Pforzheim University of Applied Sciences
Pforzheim, Germany

Florian Rupp
Kutaisi International University
Kutaisi, Georgia

ISBN 978-3-031-23014-1 ISBN 978-3-031-23015-8 (eBook)
https://doi.org/10.1007/978-3-031-23015-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The present volume is a selection of articles in honor of Manfred J. Holler's 75th birthday. As researcher, he contributed remarkable insights to various fields over the many years he devoted himself to the study of power and responsibility. He is not only author and co-author of many articles and books on game theory, including one of the best-selling textbooks in German on the topic ("Einführung in die Spieltheorie", Springer), he also worked extensively on questions of labor economics, politics, and philosophy. Being not only a very well-educated person in scholarly matters but also a lively, entertaining, and kind person, it is no wonder that he found many collaborators and friends from various academic communities, but also among journalists, artists, film-makers, and many other professional fields.

As editors, this confronted us with a more than daring task: How to select contributors to an edited volume among such a variety of inspiring people? We exploited the fact that Manfred remains an active researcher even after his retirement. As organizer of a research seminar, the Adam Smith Seminar, he still actively operates an open research network. Our best idea was to contact contributors of this research seminar over the past few years. We know this choice immediately excluded many of Manfred's longest and closest friends, to whom we would like to apologize. But we simply saw no other way to handle such an extensive list of potential contributors without overstraining the kind support of the publisher.

The contributions fall into roughly four sub-disciplines: voting and voting power, public economics and politics, economics, and philosophy, as well as labor economics. As in the case of Manfred's own writings, several authors are highly interdisciplinary, making the classification difficult.

We owe a particular "thank you" to all reviewers, who gave valuable comments and suggestions for improvements. Without your help, this volume would not have been possible. Finally, and most importantly, we would like to thank Manfred; he has been a source of inspiration to all, and active supporter of many of us.

Untermeitingen, Germany — Florian Rupp
Hamburg, Germany — Martin A. Leroch
September 2022

Contents

Manfred on His 75th Birthday

Heinz D. Kurz

1

Dear Manfred, Barbara, friends and colleagues,

Manfred turned 75 on July 25th. We are celebrating his birthday today. I am happy to comply with the request to say a few words about the friend on this occasion. The invitation from Florian Rupp and Martin Leroch refers to a "highlight of the evening". There are highlights that can be either good or bad. In that sense, the announcement can't really be wrong. Moreover, what is a highlight for some people may be the opposite for others—it depends on the position of the head and the perspective. If

The following text was given as a speech on the occasion of a virtual birthday event which took place on September 27th, 2021.

H. D. Kurz (✉)
Graz, Austria
e-mail: heinz.kurz@uni-graz.at

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_1

we also take into account—well known to Bavarians—that "laudatio" in Bavarian means "Dableck'n", deriding, then we have every reason to be wary of climaxes.

Of course, on occasions like today's, Dableck'n must not violate the principle of only bringing up praiseworthy and lovely things, that is, only a part of the truth about Manfred. Whether it is the smaller part is just as open as the question of whether the following can be said to be true at all. Not even Manfred can be sure of that, since, as David Hume knew, reason is always the slave of our passions and, as Adam Smith pointed out, our perception is distorted by our self-love.

Let me at the outset emphasize that my considerations are based on a pitifully small amount of observations about Manfred's doings and actions, so the word "evidence-based"—the analogue of "cool" amongs hooligans, which is used today in an inflationary manner in economists' circles, would, therefore, be completely inappropriate. I console myself with the fact that, according to Schumpeter, the sea of facts is infinite and mute, and which subset of facts is relevant for what? The result of studies depends, therefore, crucially on the facts selected and—no less importantly—how one makes the former speak, which theory one has at one's disposal. Applied to our case, to the perception of the "Gestalt" of Manfred, I try to portray a kind of comprehensive piece of art on the basis of just a few of its perceived salient features. Most of Manfred must necessarily remain in the dark. Who dares to say what this saves us from!

2

I should briefly state the empirical basis of my considerations. Manfred and I were born in the same year—he barely two months after me. However, we only got to know each other in the late 1960s when we were studying economics in Munich. I remember Manfred for his clever remarks and questions in lectures and seminars, but even more for his mischievous look.

What can words express compared to eyes! But as soon as we became aware of each other, Manfred disappeared together with his eyes to Lausanne to study law. Law at Léon Walras' former workplace? How good that he soon returned to Munich! If I am not mistaken, we met again in Giovanni Heinemann's exam preparation course. Manfred caught my attention again with his clever comments and questions, and of course with his eyes. In the summer semester of 1971, we obtained our diploma in Munich. Manfred then went on to study political economy. I followed Albert Jeck to Kiel as a research assistant. And so Manfred and I lost sight of each other for several years. Only coincidence brought us together again, and the coincidence I am alluding to was far more coincidental than the one described by Karl Valentin in the famous "Orchestra Rehearsal". Manfred was now a professor in Aarhus and I was one in Bremen. We met on the train from Hanover to Munich—it must have been towards the end of my time in Bremen and shortly before my time in Graz, probably in 1987. It was great to see him again.

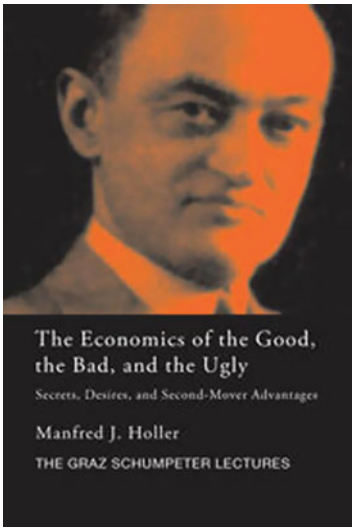
In the meantime, Manfred had lived several lives in one, including as a scientific entrepreneur, a founder of publishing houses and periodicals, an intellectual networker, a successful theorist, a textbook author and so on and so forth. His activities manifested themselves in what he said, lectured, proclaimed, founded, organized, designed, printed and published. His works are far too numerous and varied to be mentioned here. Also, there is the hope and expectation that before long entirely new tools and algorithms will be developed by Big Data aficionados and evidence-based people that will allow us to scour ether and airwaves for the voices and sounds of every human being that has ever walked and talked on planet Earth. I am already today certain of two results of this future meritorious activity: First, Manfred will be assigned a share of the scraps of words collected far above-average, and secondly, his contribution to what will one day be called anthropogenic vocal ether pollution will prove to be far below-average. In other words: while Manfred talks a lot, what he says is usually not stupid, on the contrary.

After the aforementioned meeting on the train, Manfred was kind enough to appoint me to the editorial board of the *European Journal of Political Economy*, which he had founded. I am afraid to say that I have not been of much help to him in this capacity and I hope he will be lenient. Anyway, we were in touch again.

But it was not until about twenty years later that we became much closer again, in Graz, that is. Manfred had meanwhile got to know larger parts of the world and they had got to know him. He had been a full professor in Hamburg since 1991 and had held guest professorships from Paris to Hangzhou, from Rome to Gothenburg, and from Western Ontario to Turin and Turku—impressive proof of his great international reputation and his highly esteemed personality. He had made a name for himself in numerous fields including microeconomics, game theory, economic theory of politics, labour market theory, analysis of political and economic power, electoral analysis, theory of coalitions and collective action, the question of standardization, the theory of negotiations, to name but a few![1] He had also launched other successful periodicals, including *Homo Oeconomicus*. But Manfred is not only a formidable social scientist, but he and his congenial partner Barbara do also have an extraordinary knowledge of history.

[1] The importance of his pioneering work in the field of power indices has only recently been confirmed again in publications by others. And after these lines had been written, in June 2022 he received the great honour of getting elected into the Finnish Academy of Science and Letters.

And this man, bursting with knowledge down to the tips of his hair, was a guest professor in Graz in the winter semester of 2011–2012! How nice to have been aware again that his mischievous look was resting on me! It was clear that such a talented, versatile, original and productive man had to become a Graz Schumpeter Lecturer. And so it happened in 2013. The topic of Manfred's related lectures was "The Economics of the Good, the Bad, and the Ugly: Desires, Secrets, and Second-Mover Advantages". The lectures were then published in the *Graz Schumpeter Lectures* series (*GSL*) by Routledge, London, in 2019.



Before Manfred began his lectures in Graz, he showed some important scenes from Sergio Leone's film, starring Clint Eastwood, Lee van Clef and Eddie Wallach. I had seen the strip for the first time in the 1970s in Ouagadougou, today Burkina Faso, on the occasion of a visit of my later wife Gabriele, who worked there in an agrostological project. The French title of the movie is "Le Bon, la Brute et Le Truand". It was shown in an open-air cinema.[2] The translation of "Draw your revolver!" in French is "Tire ton pistolet". Could this signal deadly danger and can serious science be done about it? The tsetse flies carrying sleeping sickness buzzing around me in Ouagadougou were a real threat, I thought. But of course, game theorists had picked up the topic of the movie and Manfred, equipped with a huge arsenal of game theory and sophisticated analytical tools offered by economic and political science, sought to deepen the answers given to the questions raised. There was no room for the sleeping sickness that is rampant in lecture halls. In his treatise, Manfred corrects and enriches Schumpeter's theory of innovation. He illuminates the concept "desires", and shows why it is often not the heroes of Schumpeter's story, the "pioneering entrepreneurs", who are successful, but those who follow them, the second movers. People and organizations often have a desire for secrecy, among other things in order to secure second mover advantages for themselves or to circumvent those of others. We owe Manfred a great painting of the conflicts, social problems and failures of politics and companies that result from the interplay of desires, secrets and second mover advantages. But he also has good news, showing that some of the problems that arise hold themselves the key to a solution.

After this highly personal reminiscence concerning the history of our encounters, now something objective, that is, something that goes beyond purely subjective

[2] It deserves to be noted that in French the word for tyrant is feminine: *la brute*. As we know, language reflects peoples' experiences across centuries.

impressions and based on unreliable memories of times long past, in short: now something evidence-based, factual.

3

Manfred's actual birthday, July 25, is the nationwide day of the sea rescuers in Germany, in the USA it is several things at once: the National Carousel or Merry-Go-Round Day, the National Thread the Needle Day, the National Culinary Day, to Christians the day of honour of the apostle Jacob, and worldwide the Day of Joy.

But what has this evidence to do with Manfred? Nothing, on the surface, everything, in essence! That the Day of Joy fits perfectly with Manfred needs no further explanation. But are there also connections in the other cases? The world is as you imagine and react to it. Or does someone seriously want to claim that a Hamburg university professor can never loose the thread? Just look at some of his colleagues in the Hanseatic city to get an idea of what I mean. Or does anyone want to deny that Manfred could ever become nauseous in the always accelerating merry-go-round of life and show signs of madness? And who would seriously doubt that even the most patient of all patients, Manfred, repeatedly lost his patience while threading the needle?

It is a good thing that July 25th is also National Culinary Day. Manfred soothes his mood there, preferably with roast pork, dumplings, coleslaw and wheat beer. The question: Why not celebrate this every day? The clever one has long since decided this for his part.

But what is the relationship between Manfred and Jacob? We can probably rule out that he will have a fate as bad as the one this apostle had according to the legend. But a parallel comes to one's mind: just as Jacob is said to have contributed to driving the "infidels" out of Spain, Manfred is helping to eliminate ignorance. Manfred, we can say without exaggeration, is on a campaign against superstition, misconceptions and misunderstandings—or, to put it in Bavarian terms: against nonsense. If it happens to be true that human stupidity is greater than the expansion of the universe, then Manfred has taken on the greatest of all challenges.

Will he prevail? Of course, we wish him to do so from the bottom of our hearts, but we cannot help recalling the words of the great Voltaire, who was convinced that the world is no better when we leave it than when we entered it. "No better" does not exclude, of course, "significantly worse". In view of the numerous kinds of madness experienced these days, one gets the impression that things are going downhill at an accelerating speed. Man has developed techniques, which allowed him to conquer and subjugate the planet. But as in so many other cases too, the danger is that once you have painstakingly learned to master a technique, it will take control of you. Anyone who knows how to use a hammer sees nails everywhere. Anyone who has learned optimizing techniques often ends up in the imaginary world of Dr. Pangloss in Voltaire's *Candide*, in which everything is assumed to be in the best order possible. However, what is penny-wise may turn out to be pound-foolish.

There is an unmistakable measure of Manfred's considerable success as an apostle of reason—his publications. Someone who has written and published more than 70 books alone or together with others in about 50 years of study and work, some together with Barbara and others with several colleagues who are virtually present today, with some books going through several editions; someone who has published well over 100 articles in academic journals, alone or together with others; someone who has placed well over 120 contributions in anthologies and has also written numerous book reviews; someone like him is truly what is called a "Kapazunder" in Austria

or a luminary in the English-speaking world. I imagine the days and nights full of deprivation at the desk and in front of the computer that Manfred sacrificed for the progress of science. Manfred has wrested the remarkable results of his work by using all his mental and physical strength, repeatedly completely exhausted and with red eyes from overly long seances in front of the screen of his computer, only at the first cockcrow did he sink onto the bed, when many of us were still comfortably lounging in the sheets.

But Manfred is not only an author, as already mentioned, he is also the founder and editor of several specialist journals and book series. As has already been mentioned, he wrote several of his works with Barbara, most recently game theory for managers.[3]

In addition, he is the director and/or member of several research centres and projects near and far, including one on conflict resolution. Machiavelli's *Il Principe* has been on his mind for a long time. Manfred effortlessly transcends narrow disciplinary boundaries, overcoming the limitations of disciplinary thinking without jeopardizing intellectual discipline, reaching out to philosophy, literature, history and art—together with Barbara and his friends and students as complementary think tanks. Manfred's thoughts and aspirations are based on the example of the *homo universalis* of the Renaissance—an ideal that is unattainable today. But it is impressive how far Manfred is getting along the way. No question: Without Manfred and his work, the world would be different. His CO_2 balance is also remarkable, indeed commendable, as he does not own a vehicle with an internal combustion engine. He

[3] In the late 1980s, the editors of *The New Palgrave* were amazed at the huge sales of the expensive four volumes in Japan. Their attempt to trace the reasons for this revealed that Japanese managers had purchased the volumes for their secretariats in bulk. They had apparently interpreted the subtitle "A Dictionary of Economics" as indicating a handbook useful to their language secretaries in their correspondence with English-speaking countries.

and Barbara cover distances on foot, by bike or by public transport. Well, every now and then they use an aeroplane.

Looking at all his achievements, one feels small as a colleague and even more so as a laudator who, by virtue of his office, has to get an overview of the scientific achievements of the honouree. More precisely: every additional publication by Manfred one looks at, tends to make one shrink a little more. When you accept the office, you naturally feel a friendly bond with Manfred and consider the task honourable. But this feeling was put to the test in the course of my preparations for tonight. How can Manfred so ruthlessly expose his friends to the swelling feeling of their own mediocrity! Manfred is well acquainted with the Scottish Enlightenment and the doctrine of the unintended consequences of one's own actions and omissions. It can therefore be assumed that he is aware of the sense of failure that his successes prompt in his admirers. The question is close at hand: is it not so much about unintended but about consciously intended consequences? Does he intentionally make us feel our own puniness? Is Manfred, as the Bavarian would put it, malicious "with diligence"? Is his look mocking and condescending rather than mischievous?

We have gathered to praise Manfred and to assure him of our appreciation, affection and friendship. But do we own his? Or are we dealing with a male mutation of the "old moor witch" of a German fairy tale, about whom a nursery rhyme says: "Hält die ganze Welt für dumm, hext herum, hext herum" (Considers the whole world as stupid, witches around, witches around).

Manfred, honestly: Do you think what the witch says? In order to really do you justice, we need to know how you feel about us. If you walk ahead of us at a brisk pace, do you do this in order to be able to mumble unheard unflattering judgments about us?

When gurgling wheat beer jumps down your throat, do the bubbles transport disrespectful things about us? Even worse: Does the arrangement of letters in your writings contain secret messages which, when deciphered, say, for example: "Heinz is dumb" followed by: "and he doesn't even notice it"?

The more I think about it, the more I get lost. Manfred, dispel the fog! Remember the philanthropic core of Ricardo's theorem of comparative advantage, which, mind you, contains the good news: Each one of us may be inferior to you in each and every respect, but your superiority varies in degree. So there are mutually beneficial cooperative relationships out there. You too benefit from relationships with us, not just us from relationships with you! Provided, of course, that each one of us specializes in full recognition of your different relative degrees of superiority with regard to the various dimensions under consideration.

It is no coincidence that you are honoured today. The fact that the task of honouring you fell upon me, is much more difficult to explain and is best seen as a whim of fate. But imagine that for whichever reason our roles would be reversed? A quick reflection shows that this would mercilessly reveal the limits of Ricardo's theorem and, I am afraid to say, potentially also the limits of your relationship with us. Or am I mistaken? The question I am asking, having learned important lessons from you, is this: Did you move too soon and jump ahead too early, thereby pushing each one of us into the position of a potential second mover with a bright future ahead of him or her? Does your own theoretical reasoning in the *GSL* in the end turn against you?

I like to flee into such fantasy worlds, but they quickly burst vis-à-vis your untouchable superiority. What remains to us is to wish you belatedly a happy 75th birthday. We do so aloud with the words of the people of Holofernes. According to Nestroy, this people chants in his play *Judith and Holofernes* towards the end of the first scene:

"Weil er uns sonst niederhaut,

Preisen wir ihn Alle laut!”

(Because otherwise he’ll cut us down,

Let us all praise him loudly!).

While, as has been stressed at the beginning, the whole Manfred remains hidden from us willy-nilly, we can rule out with a probability bordering on certainty that he will have a fate like that of Holofernes. So: Manfred, keep your head and remain as you are! Or if, surprisingly, you should not be able to do so, try to become an even better being than the one you are and fight the decline of humankind.

To you and Barbara:

Ad multos annos!

Economics and Philosophy

Three Types of Dramatic Irony

Timo Airaksinen

1 The Definitions of Dramatic Irony

Peter Goldie provides two examples of dramatic irony:

> In Shakespeare's *King Lear*, there is a scene that involves a very powerful use of dramatic irony. Gloucester, who has recently been cruelly blinded, wants to die. He asks Edgar to take him to the "very brim" of the cliffs of Dover, to "a cliff whose high and pending head / Looks fearfully in the confined deep" (Act IV Scene i). Edgar misleads him into thinking that he has done just that. […] The audience knows that what Gloucester does not know: that, contrary to what he thinks, he is not on the edge of the cliffs of Dover, and thus not able with one step to cast himself over the edge to his certain death. This is dramatic irony.[1]

Compare this with Goldie's second example: "[I]n Sophocles' *Oedipus Rex*, the audience *knows* that Oedipus killed his father at the crossroads, but Oedipus *thinks* he killed a stranger."[2] *Macmillan Dictionary* agrees: dramatic irony occurs in "a situation in which an audience *knows* more about what is happening in a play or film than the characters do." *Britannica* adds to this: "the words and actions of the characters […] take on a different—often contradictory—*meaning* for the audience than they have for the work's characters." Richard Nordquist mentions both the idea of different meanings and the doxastic gap.[3] *Dictionary.com* says: Dramatic irony entails "*irony* that is inherent in speeches or a situation of a drama and is understood by the audience but not grasped by the characters in the play" (my italics).

[1] Goldie (2014), p. 27.

[2] Goldie (2014), pp. 26–27.

[3] Nordquist (2020).

T. Airaksinen (✉)
Faculty of Social Sciences, Department of Practical Philosophy, University of Helsinki, 00014 Helsingin yliopisto, Helsinki, Finland
e-mail: timo.airaksinen@helsinki.fi

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_2

These definitions focus on (1) a doxastic asymmetry between the characters on stage and the audience, (2) the characters' interpretation of relevant situational meanings, and (3) the characters' missed ironies on stage. How are these three definitions related? We also need an intuitive notion of dramatic irony: what happens on stage and how the audience understands it is somehow problematic and unsettling. Hence, the audience finds the situation ironic. Therefore, the theory of dramatic irony must explain the emerging irony. I argue that (3) entails (2) and (2) entails (1), but not vice versa. Dramatic irony always displays (1), which may not explain or exemplify situational irony, unlike (2) and especially (3). In other words, when the audience recognizes (1), they must use (2) and (3) to explain the resulting ironies. But they cannot use (2) and (3) without (1).

In what follows, I understand irony in the usual manner.[4] Verbal irony entails a discrepancy between a speaker's surface meaning and real meaning, or what she says and does not say, given that the context is not metaphoric.[5] A big and robust person begs a little man not to hurt him. What he says (surface meaning) does not fit what the audience thinks he should say (real meaning). This is *verbal irony* that the speaker freely creates.[6] It emerges via describing a given social situation in a twisted manner, regardless of the situation. The speaker *makes* it look strange. *Situational irony* is what one *finds* in a twisted social situation. The example above turns into situational irony when we refuse to understand the big man's utterance as verbal irony; we think he is afraid of the little man. If the big man is not afraid when he speaks, the irony is of the verbal kind; if he is afraid, this is situational irony. A minimalist criterion of an utterance or a social situation being ironic follows: the case entails irony if and only if we can detect two mutually inconsistent but related interpretations of it. All three definitions of dramatic irony satisfy this criterion: the audience and the characters see the case on stage differently. The question to ask is, when is the issue significant enough to earn the epithet "dramatic irony"?

Dramatic irony posits characters on stage, a scripted scene, and an audience. We assume that the audience can perceive, understand, and interpret the scene on stage. Sometimes the relevant cues are subtle and ambiguous, and the audience members may disagree and perceive the scene differently. Therefore, we must assume an intelligent, well-informed, and attentive audience, which is not always the case. It has its idealized features. The audience is an idealized doxastic agent who does not miss obvious information as the characters do. The real audience may miss and misread information, unlike the ideal audience that reliably acquires the relevant beliefs. We are interested in an ideal audience when discussing dramatic irony. When we focus on real audiences, they may miss the ironies of a scene. In this case, ironies go unnoticed, but the unnoticed irony is not irony.[7] The idea of a stage may be *als*

[4] See my (2020a) and (2020b).

[5] Such an elementary definition does not distinguish between irony and metaphor; see Dynel (2016).

[6] Richard Rorty seems to think that any situation can be read ironically by an "ironist." See Rorty (1989), p. 73. Also Inkpin (2013), and my (2021a).

[7] The idea of audience is problematic here, in rhetoric, and in argumentation theory. Perelman and Olbrechts-Tyteca (1971) introduced universal audience that "provides shared standards of

ob. Sometimes the characters are on a theater stage, but they may also act in other stage-like situations. The roles and scenes follow a script, or we can read them as if they did: without the script, no characters exist. Characters exist because the script defines them, and they act on stage in front of an audience.

2 The Limits of Doxastic Theory

Dramatic irony may entail a doxastic asymmetry between the audience and the dramatic characters on stage. The audience knows more than the characters who stay ignorant, mistaken, or misguided. The doxastic supremacy of the audience comes about in two ways. The audience may be independently informed, or they acquire their beliefs from the unfolding events on stage. Suppose they read the script beforehand and then see the play. Or they see the play many times over. Like a powerful divinity, they now possess complete knowledge of future events. Does this mean they must now watch the whole play ironically? Everything that happens in every scene is now supposed to be ironic, which is hardly true.

Moreover, the audience will lose the enjoyment based on the naïve viewing that permits immersion into the events as they progress on stage. We may not want to adopt such an independently informed standpoint, even if we could. The doxastic theory looks too general to be viable. It finds irony where it may not exist.

Popular artistic conventions create scenes that look dramatically ironic in the doxastic sense when they are not. They are evident in popular cultures, such as Hollywood films. Conventions indeed offer the audience a chance to ironize their experiences, but this is not dramatic irony, as I will show. Classic Westerns, even the best ones, are so conventional that they may look ludicrous today. An example is Fred Zinneman's *High Noon* (1952). Bad men arrive in town, but the audience realizes they must die as scary as they may look. The contrast between their appearance and predictable fate is ridiculous—it is pure comedy. As the audience predicts, the good guy, sheriff Will Kane, will win the ensuing gun battle. If a gunshot hit him, the wound should be on the shoulder. And bullet wounds are never too painful. They extract the bullet, give him some whiskey, and the wounded man is healthy again. The audience knows all this beforehand. Yet, they cannot predict *how* the bad men die, and thus the viewing is stress-free and enjoyable. This type of foreknowledge aims at stress reduction.

Conventions also rule Hollywood sex. After making love, the woman wraps herself in white bedsheets to her neck, and her bare-chested man used to smoke a cigarette (it is no longer permitted). The audience may skip these conventional ironies because they are so familiar. Technically, such scenes exemplify doxastic dramatic irony, but

---

agreement by which to measure argumentation"—and we can add, the level of observation (p. 133). See Tindale (2004), p. 133. Aikin tries to save the much-criticized idea that is much too vague (Aikin, 2008). See also my (2022), where the idea of audience is central.

only in a trivial sense. Films need their conventions, which entail no irony in any interesting sense. We do not want to force irony on everything.

Will Kane is not Will Kane but Gary Cooper. The audience knows this, but this hardly qualifies as an example of dramatic irony. The case of John Wayne is subtler: he always plays John Wayne even when he plays Ethan Edwards in *The Searchers* (1956), and of course, sheriff John T. Chance in *Rio Bravo* (1959). Isn't it dramatically ironic that John Wayne is John T. Chance, who is John Wayne? John Wayne does not act out because he is who he is on stage? He always is the same invincible hero. The audience may realize this as situational irony, but it hardly qualifies as dramatic irony.

The audience often witnesses events on stage of which some characters remain ignorant. Alfred Hitchcock's *Rope* (1948) is an example. The audience sees a murder of which some characters initially know nothing. Rupert (James Stewart) solves the case, and then they all know. Here the audience possesses information that is internal to the script. An independently and externally informed audience reads the script beforehand or leans on the relevant conventions. The Gloucester case exemplifies internal information because the audience knows what happens simply by watching the play.

Sometimes a character knows more than the audience. In Kafka's *The Castle*, K. knows, contrary to what he says, that he is not a land surveyor—a naïve reader may miss this. It is not easy to see through K.'s lies.[8] Should we call the beginning of *The Castle* dramatically ironic because of its inverted doxastic asymmetry? K. knows more than the audience. Agatha Christie's character Hercule Poirot always knows the name of the murderer before the audience does—and the audience wants to be surprised by him. Is this dramatic irony according to the doxastic theory? Hercule Poirot and the murderer cannot know they know more than their audience, for obvious reasons. The audience knows that they know less, and they try to overcome this deficiency through guesswork and logical reasoning. This inverted case is dramatically ironic in the doxastic sense if the only thing that matters is the doxastic asymmetry and not its direction. We may call such a case an extension of dramatic irony.

What about Kafka's K.? The case is obviously ironic. Perhaps the ambiguity of K.'s position is the key: he insists he is a land surveyor and invited by the castle, but a critical audience need not accept this. The castle does not recognize him, and most importantly, he can find no way to enter there. What we see here exemplifies situational irony and, perhaps dramatic irony, if possible, in a stronger sense than in the case of Hercule Poirot. K.'s situation is radically less convention-bound than Poirot's—K-'s lies are hard to detect, and in the end, the case remains controversial and unsettling.

In some cases, the inverted doxastic asymmetry indeed creates dramatic irony, at least intuitively. The inverted asymmetry is dramatically meaningful to the audience. The script ridicules the audience, who cannot say whether K. tells a lie. Or he may be honest, but the audience has a hard time deciding. All this entails dramatic irony in the internal but non-standard sense.

[8] Kafka (2019). See Steinberg (1965), and my (2017)

The following key example shows that the idea of the dramatic irony of doxastic cases is too general to be viable as a definition. Suppose we have a scene on stage where a person sits in a large armchair facing away from the door. The door is open, and two persons emerge at the door. They discuss something, and the third person hears them without revealing her presence in the room. In such a situation, the audience knows more than the discussants, namely, that a third person is in the room and can hear all they say. The discussants disappear from the door. This example works equally well when the person in the armchair does not know about the two persons at the doorway. The audience learns about them, unlike the third person. As we'll see, the problem is the meaninglessness of the knowledge here.

First, let us suppose the discussion was trivial and did not interest the third person; second, the conversation was relevant to her—perhaps it was insulting. According to the doxastic theory, the first case is ironic. It is not. The trivial fact that the characters did not know about each other is meaningless. Suppose the discussion was meaningful to the person in the armchair who heard it. The dramatic irony follows because of the doxastic gap and the meaningfulness of the issue. Someone will be in trouble later on, as the audience now knows. The scene was meaningful, but unlike the audience, the characters do not know it yet.

If the topic was of interest to the third person and she heard it, the case is dramatically ironic. However, the doxastic theory cannot distinguish between the first and second versions. This theory only talks about beliefs and knowledge, but in the first and second cases, these remain the same: the third person is in the room, and the audience, unlike the two discussants, knows it. But the two cases differ in the meaningfulness of the discussion at the door. The dramatic irony requires that the third person hears something she should not hear. Unlike the discussants, the audience can see how the play's plot depends on this event. The third person is now meaningful to the subsequent events, which the characters at the door fail to anticipate.[9]

This simple argument shows that the doxastic theory alone fails to explain dramatic irony. The idea of dramatic irony depends on the meaningfulness of the events on stage, but it also presupposes doxastic asymmetries. The audience knows more of the scene's features that are meaningful to the characters—the script makes them unable to realize it.

3 The Meaning Theory of Dramatic Irony: Sophocles' *Ajax*

The audience and the characters on stage sometimes attach a different meaning to dramatic events so that the characters will miss something. This is different from the doxastic theory. When Oedipus kills his father at the crossroads, he does not know what he is doing. He knows that he kills but not whom he kills. What is the meaning, for him, of this critical event? It is unimportant, but knowledgeable audiences know better. Their beliefs are accurate. Think of Sophocles' *Ajax*, where the great warrior

[9] I am grateful to Dr. Heta Gylling for the basic idea of this example.

Ajax falls raging mad and kills a flock of sheep that he thinks are enemy soldiers.[10] He initially assigns a significant but mistaken meaning to his act: he sees valor in it. He kills himself when he realizes what happened and understands the situational irony and the sarcastic import of the event. What else could he do when he faces his comrades in arms?

The scene is situationally ironic because of Ajax's interpretation; the surface meaning differs from the true meaning. The surface meaning is heroic, but the real meaning is ridiculous and humiliating. He deserves the ridicule. Here the situational irony transforms into dramatic irony in front of a knowledgeable audience. A brave hero killing sheep sounds ludicrous, which the audience realizes well before the hero comes to his senses and understands the real meaning of his actions. Here the ancient moral context is different from the modern one. We think Ajax's madness and Oedipus' ignorance excuse them, which is not what the ancient audiences believed. Ajax and Oedipus are guilty, and both punish themselves cruelly.

Tecmessa, Ajax's concubine, explains the scene and the different perspectives of Ajax and his audience. Ajax fails to interpret his actions correctly, and his madness explains this:

> TECMESSA: Yonder man, while his spirit was diseased, / Himself had joy in his own evil plight, / Though to us, who were sane, he brought distress. / But now, since he has respite from his plague, / He with sore grief is utterly cast down, / And we likewise, no less than heretofore. / Are there not here two woes instead of one?

Indeed, we recognize "two woes instead of one" in this scene. The audience saw the truth where the character could not.

The Ajax case also exemplifies the doxastic theory. Ajax does not know what he is doing, unlike the audience. But the more profound point focuses on Ajax's pride in what he did. A great warrior attacking sheep is paradigmatically ironic—a sheep is a metaphor for defenseless vulnerability. When Ajax finally understands what happened, he commits suicide. The relevant ironies are there to explain the events. The scene's logic depends on Ajax's misreading of the facts, that is, of the false meaning he assigned to them. He does not commit suicide because now he knows what he did—he killed sheep; he must die because he knows the *meaning* of his action. His behavior was ridiculous. And the audience understands all this before Ajax does: a sheep killer warrior is an oxymoron. Ajax's new identity as a warrior and fool is hopelessly confused and no longer allows a consistent description. We can read the sheep killing scene according to the second theory of dramatic irony.

[10] Sophocles (1919) This translation does not provide line numbers.

4 The Irony of Missed Ironies in Euripides

In his oft-quoted article, Gareth Williams refers to "the privileged position of the reader of *Heroides 11* who, through access to the *Odyssey*, is alive to ironies which Ovid's Penelope cannot realize." Hence, he says, we find the third type of dramatic irony here.[11] The critical point is the missed situational irony that turns into dramatic irony when the audience understands it. But Williams does not hold this position self-consciously or consistently, as shown when he discusses the following example:

> She is right to equate Aeolus with the winds, but she does not know enough to appreciate the full force and accuracy of the comparison. Her father's change of heart enables the privileged reader to realize the full potential of the comparison which is impossible for Canace herself.[12]

Along with the idea of "accuracy of comparison," he may return to the doxastic theory of dramatic irony. This wavering is typical when one works with an unanalyzed notion of dramatic irony. And the following hints at the priority of the meaning theory to the doxastic one: "The true comparison between Aeolus and the winds now lies not in their shared ferocity, but in their common changeability."[13] The meanings of the relevant metaphors have changed, but Canace, unlike the audience, does not realize it.

Euripides' tragedy *Bacchae* provides examples of the dramatic irony of the third kind. It also illustrates verbal irony in dramatic settings. The characters miss ironies *de se* on stage, that is, ironies concerning them personally.[14] Dionysus, also called Bacchus and Bromius, arrives at Thebes as a human being and announces himself to the court of king Pentheus as a god who insists on his rites.[15] The king rejects the stranger's divine status, which is a mistake and leads to tragic consequences. The audience knows that Dionysus is a godhead born to Zeus and Semele, a human woman. The audience also knows the conventions of tragedy: Pentheus must perish together with his house. They realize that Pentheus should understand and yield to the god.

Bacchants are already reveling in the nearby hills and woods, but this does not convince Pentheus; it makes him curious, and he wants to see them. Pentheus should and could have known better. Therefore, his royal arrogance is misplaced, and Dionysius mocks him in a threatening manner, promising him a quick death. For him, one's name is an omen, or *Nomen est omen*:

> DIONYSUS: You're quite ignorant of why you live, what you do, and who you are.
>
> PENTHEUS: I am Pentheus, son of Agave and Echion.

---

[11] Williams (1992), p. 201. Tragic Irony is just another word for dramatic irony; see Casali (1995), p. 509f.— Huson (1998) argues that, according to Hegel, "[t]ragic irony is constituted by the self-destruction of a historical subject in whose downfall a *higher objective value is revealed*." (p. 123). This idea is not related to dramatic irony. Also Phillips (2009).

[12] Williams (1992), p. 209.

[13] Idem.

[14] *De se* means "concerning the person". See Torre (2016).

[15] Euripides (2017). See Segal (1997).

> DIONYSUS: A suitable name. It suggests misfortune.
>
> (Lines 630–640)

This sarcasm turns into situational irony when Pentheus threatens to punish the god as if a mortal human being could do that—again, the audience knows better. Dionysus puts the situational irony into words:

> DIONYSUS: What punishment am I to suffer? What harsh penalties will you inflict? (615–616)

The situational irony is evident when the god dresses the king in women's clothes to smuggle him to the orgies of the female Bacchants. But before this happens, the citizens of the *polis* can see their king in a humiliating position dressed as a woman. Again, the audience knows more than Pentheus, which entails dramatic irony. But most importantly, Pentheus misses the situational irony: someone leads the king through his city wearing women's clothes. The audience cannot ignore this irony: the surface is a person walking through the city, and the real concern is a humiliated king. The walk is not what it looks, but Pentheus fails to see its ironies.

An additional source of irony is that the early Christian tradition sometimes confused Jesus Christ and Dionysus.[16] Both are upstart Eastern gods, both born of a woman conceived by a godhead, and they assume the human form, yet insisting on their novel rites. Moreover, their identity is a trinity. Both have three names, and in this sense, they are polymorphic beings, or they are one person in different simultaneous guises.[17] Their names form two triune metonymic groups: Dionysus, Bacchus, and Bromius; God, Spirit, and Son.

Jesus, like Dionysus, uses verbal irony and sarcasm, for instance:

> The Jews picked up stones again to stone Him. Jesus answered them, "I showed you many good works from the Father; for which of them are you stoning Me?" (John 10:31–32)
>
> Just at that time some Pharisees approached, saying to Him, "Go away, leave here, for Herod wants to kill You." And He said to them, "Go and tell that fox, 'Behold, I cast out demons and perform cures today and tomorrow, and the third day I reach My goal.' Nevertheless I must journey on today and tomorrow and the next day; for it cannot be that a prophet would perish outside of Jerusalem." (Luke 13:33)

These exchanges emphasize that the audience should know better, that is, Jesus indeed is a divine being. He derides them: you do not believe in me although you should—and in the end, you must! The point is: you threaten to kill me, but instead, you will kill yourself. His irony is a warning of a threatening tragedy. This entails the third kind of dramatic irony: both Jesus and Dionysus speak like gods, or sons of a god, although their listeners do not know and understand it, unlike their audiences. Their listeners may not appreciate these ironies, but to miss irony is ironic.

As we see, in addition to verbal play, the Biblical narrative contains dramatic irony. Jesus has his audiences, the readers of the Bible and his contemporary listeners to

---

[16] See Friesen (2014). István Czachesz (2014) argues that Christ was confused with Apollo, Asclepius, Dionysus, Hercules, and Helios etc. (p. 212).

[17] Czachesz (2014), Ch. 7.

whom he announces his divinity, promising life to the believers and sinners death.[18] This is what Dionysus, that cruel Eastern upstart godhead, says, too. Jesus may have provided his contemporary audiences sufficient evidence of his true nature and offered them a chance to show their devotion, yet they failed. The characters in this great religious drama should have known better, which is the foundation of the situational irony that culminates in dramatic irony.

Notice an additional similarity between Jesus and Dionysus: they are ambiguous figures, human beings, and gods. Dionysus says:

> DIONYSUS: Yes, I've changed my form from god to human, / appearing here at these streams of Dirce, / the waters of Ismarus. I see my mother's tomb— / for she was wiped out by that lightning bolt. / It's there, by the palace, with that rubble, / the remnants of her house, still smoldering / from Zeus's living fire—Hera's undying outrage / against my mother. (5–12)

He says he is godly, this is his assertion of identity, yet in other places, he says the god sent him to act as his messenger. Both Jesus and Dionysus are both human and divine.

The king and the god discuss, Pentheus trying to be ironic:

> DIONYSUS: I'm from there. My home land is Lydia.
>
> PENTHEUS: Why do you bring these rituals to Greece?
>
> DIONYSUS: Dionysus sent me—the son of Zeus.
>
> PENTHEUS: Is there some Zeus there who creates new gods?
>
> DIONYSUS: No. It's the same Zeus who wed Semele right here.
>
> (570–580)

Dionysus is indeed a god, but the person Pentheus is now addressing is no longer identical to the god. He is a messenger, although he has divine powers. Jesus was like this, but his final words betray him: "My God, My God, why have You forsaken Me?" (Matthew 27:46; Mark 15:34). The question mark here is problematic—the speech act is an accusation that forces dramatic irony into the situation: suddenly, he fails to see what his audience is seeing.

The similarities between Jesus and Dionysus are evident. We learn that their listeners should believe in them, offer them their rites, and worship them—the alternative is death. In both dramas, human characters should know that they must suffer now and forever if they fail. Both plays are cruel because the script dictates that their audiences do not know and believe, although they should. The script condemns them. Should we feel a genuine temptation to warn the characters who listen to these two cruel Eastern gods, namely, "Please, have faith in these two; they are going to humiliate and kill you"? They fail, and this is where the dramatic irony starts: unlike

[18] Kenneth Burke (1970) writes: "The Bible […] teaches us that tragedy is ever in the offering. […] Let us be on guard ever, as regards the subtleties of sacrifice, in the fundamental relationship to governance" (p. 235).

the ideal audience, the characters miss the situational irony inherent in mistreating gods.

When Dionysus leads Pentheus in woman's clothes through the streets of Thebes to his doom in the hands of the raving Bacchae, the audience feels not only that Pentheus should know, but he could know better. His arrogance costs him dearly in the hands of the god who feels Pentheus has betrayed him and caused him unduly unruly harm and pain. However, the ultimate situational irony here results from the strange vulnerability of the god: how can a humble human being cause so much damage to a noble god, or why does he so desperately need his rites? Dionysus never tells, nor does Jesus. Indeed, this entails dramatic irony. It is based on situational ironies that neither the king nor the god can see. The audience realizes that Dionysus fails to appreciate the *de se* ironies of being such a needy god. The mighty god is weak. Or, the powerful are weak without making the weak mighty. Here is another example of the third type of irony:

> DIONYSUS: You've heard what I had to say, / Pentheus, but still you're not convinced. / Though *I'm suffering badly at your hands*, / I say you shouldn't go to war against a god. / You should stay calm. Bromius will not let you / move his Bacchae from their mountains. (My italics.) (963–968)

Perhaps the god speaks ironically. Suppose he does. It does not change the ironies of being dependent on mere mortals. However, if he does *not* speak ironically, the third type of dramatic irony arises. In this case, the god misses the ironies of being a needy and vulnerable god.

5 The Two Functions of Dramatic Irony

Irony is a pragmatic trope, and thus dramatic irony affects the audience. Nordquist writes:

> The function of dramatic irony is to sustain the reader's interest, pique curiosity, and create a contrast between the situation of the characters and the episode that ultimately unfolds. This leads to the audience waiting in fear, anticipation, and hope, waiting for the moment when the character learns the truth behind the events of the story.[19]

He continues, "Readers end up sympathizing with the main characters, hence the irony."[20] How does ironic treatment create such an S-effect (sympathetic effect)? Irony typically entails alienation and a Brechtian V-effect (*Verfremdungseffekt*)[21];

[19] Nordquist (2020).

[20] This irony is not dramatic irony because the dramatic irony focuses on what happens to the characters. Nordquist (2020) focuses on what happens with the audience. It is situational irony *simpliciter*, if it is irony at all.

[21] Berthold Brecht suggested that evil characters should wear a mask or even a crocodile oufit on stage to eliminate any threatening S-effect; see Brecht (1987, p. iv). See also my (2021b). How to stop the audience from liking the evil characters like Puntila on stage? Dramatic irony may bring about the S-effect but other factors do the same.

now dramatic irony should make an S-effect, which is the opposite of alienation; why call dramatic irony "irony" if this is the case? The possibility of the S-effect alone may not refute the idea that irony entails alienation and estrangement.[22] S- and V-effects are mutually incompatible, and irony loves the V-effect. If the audience only recognizes an S-effect, they miss the ironies entailed by the V-effect. Without it, ironic games hardly are worthwhile. Think of Gloucester. The audience may miss the irony of the scene. Thus, they feel pity and sympathize with him directly. Or, they see the irony, which entails the V-effect. They indeed may also feel pity and sympathy, but as well, they may stay at the ironic level and focus on his gullibility.

The V-effect comes first, and S-effect bypasses the irony. In the Oedipus case, it is easy to miss the irony and pay attention merely to the dire consequences of his actions. Hence, the V-effect is an essential element of irony, whereas the S-effect is contingent and often irrelevant. The lesson to learn is: we must pay careful attention to the ironic effects of a dramatic scene.

The general problem of dramatic irony is this: The definition may be clear and straightforward, but audiences may have difficulties seeing the irony, especially when the S-effect is strong or the plot is too exciting and engaging. The invisible irony is an oxymoron: in a scene on stage, dramatic irony rules, but the audience may not notice it. I said above that the idea of an audience is an idealization, and an ideal audience can see and feel the irony wherever it occurs. How satisfying such an ideal solution is hard to say. Dramatic irony is a strange type of irony if the audience tends to miss it on a regular and predictable basis. To use a term borrowed from the philosophy of science, dramatic irony is then a theoretical term.

As I said, if the audience pities Gloucester, they have missed or bypassed the ironies of the case. However, the artistic value of such a scene depends on the ambiguity between its ironic and compassionate readings. The audience may react in two opposite ways. They feel the tension, and thus their experience becomes aesthetically significant. But this presupposes that the audience can first see the scene's ironies.

An ironic speech act plays with falsehoods and twisted language. Thus, it appears as prima facie dishonest. This alienates the audience from the ironic target. They must start thinking from a new perspective, as if from the outside. Also, in the case of verbal irony, the audience must ask what the speaker means. He says something he does not really mean, which creates a communication gap. The audience may try to solve the problem before reacting emotionally. The same applies to ironic situations: they all look somehow strange. But when you listen to an ironist, you may empathize with her target; you may pity him, which entails the opposite of

[22] I agree with Bennett (2016): "It is well estabished that irony is more than just saying one thing and meaning opposite [. . .] It involves what we might see as degree of detachment" (p. 234). See Dynel (2008).—Renegar and Goehring (2013) write: "[In] other words, irony allows for a both/and perspective to flourish in a world where either/or choices are often dissatisfyingly limiting" (p. 319); I call this an Eldorado View of Irony: irony gives us all the good things one may imagine—to put the point ironically.

alienation. Therefore, irony may prompt an S-effect, yet it is not irony if it initially fails to alienate its audience.[23]

Irony itself deserves a sarcastic smile, but not compassion and pity. Irony entails a V-effect but not its derivative S-effect. Situational irony may bring about, say, the feeling of vicarious shame, which indicates an S-effect. A nasty sarcasm is always verbal and may bring about well-founded empathy toward an undeserving victim. Still, as I said, this means reaction formation: the inherently alienating effect backfires.[24] An audience member may want to save the hero on stage by standing up and shouting, "Stop, please don't enter—he's got a gun." She feels the S-effect and thus misses both the V-effect and the irony of the situation. I conclude that S- and V-effects are independent of each other yet often occur together in dramatic contexts and create artistically relevant tensions. The V-effect sometimes brings about the S-effect but not necessarily. The V-effect entails alienation and S-effect togetherness.

6 The Roots of Dramatic Irony

A scene must be meaningful and engaging to create dramatic irony. Goldie's example satisfies this condition. Gloucester's situation is miserable, yet Edgar is cruelly deceiving him; thus, the audience cares. And the case is situationally ironic: the audience knows Edgar's help is a travesty. The audience may sympathize with Gloucester and condemn Edgar or laugh with Edgar, which entails a cynical attitude toward Gloucester's predicament.

The scene is charged. What is the explanation? The audience does not merely record the facts but ponders their alternatives. They see something going wrong, but what are Gloucester's alternatives? Can they say that Gloucester *should* know what is happening to him? Indeed, he should. This is the *practical should of exhortation*, which expresses the need to react and act in a particular manner. He should know and act; otherwise, he must suffer from humiliation and ridicule. Hence, he should respond: it would be beneficial for Gloucester to react now.

Suppose a teacher says to her pupil: you have already read the book; you *should* know the answer. Here is a new sense of should: Gloucester should realize the deception because he has the relevant evidence. His current location cannot feel like the cliffs of Dover. We can call this an *epistemic should*.

The should of exhortation is *practical* as it focuses on something we do and achieve:

> You should know the route, otherwise you cannot find your way home.
> You should act, or otherwise you miss a good opportunity.

[23] Golding (2012) realizes that his idea of dramatic irony is a strange one and "oddly out of place here" (p. 27).

[24] See the examples in my (2020b): Mocking, or evil mimesis, is a good example because it may create first alienation and then reactive sympathy towards the victim. Sarcasm works in the same way.

The *epistemic* should concerns evidence-based inferences:

You should know, as you possess all the relevant evidence.
As a mathematician, you should understand this proof of the theorem.

Certain situational irony in *King Lear* follows from Edgar's clumsy attempt to mislead Gloucester, as the audience must realize. Think of a person on a high ledge over the ocean and its unique sensory offerings, the sound and echo of seabirds and waves, the sweet smell of sea salt and rotten seaweed, and that whirling, moist ocean wind. The scene may work on stage, but at the same time, it fails to feel convincing. One finds much too much theater here. Does it justify the suggestion that Gloucester *epistemically should* discover Edgard's deception? If he should, this scene exemplifies the second type of dramatic irony. We can see the ironic meaning of the situation emerging when we think about its epistemic should.

Suppose the audience knows the relevant conventions that define the scene. In that case, they also know more than the characters who cannot know them, and therefore any "should know" becomes irrelevant. The characters cannot know the relevant conventions, and consequently, no "should" applies here—the scene may not allow dramatic irony. Therefore, to ask "should they know" is essential. Suppose the characters should know, and the idea of dramatic irony emerges. The audience may now genuinely expect something from the characters and not only watch them acting. We also can distinguish between internal and external dramatic irony in a novel way. The first requires that the characters can and should know. The second treats this as an irrelevant requirement. *High Noon* represents the external and Gloucester vs. Edgar the internal theory. Will Kane cannot know what will happen, but Gloucester can and should know where he is and what to do. Will Kane possesses no evidence for his final winning position. Gloucester has evidence for seeing through Edgar's evil plan. It makes no sense to say Will Kane should know (practically), unlike Gloucester, who should know to avoid humiliation.

Plato says noble lies are permissible: for instance, the lies of the prince and the doctor.[25] They may be necessary because, by lying, one can avoid significant harm. Such lies are examples of something one should not know—that is why the lies are prudentially justifiable. Machiavelli, of course, is the master of this black art. Think of this princely display of deception and cruelty, which anyhow is a justifiable political move:

> And because he knew that the past severity had caused some hatred against himself, so, to clear himself in the minds of the people, and gain them entirely to himself, he desired to show that, if any cruelty had been practiced, it had not originated with him, but in the natural sternness of the minister. Under this pretense he took Ramiro, and one morning caused him to be executed and left on the piazza at Cesena with the block and a bloody knife at his side. The barbarity of this spectacle caused the people to be at once satisfied and dismayed.[26]

The citizens praise the prince; they do not know better—although they could and should. The survival of the city state requires regal deception; without it, the prince's

[25] See Plato (2007), Book 3, 414e–15c. But see Morrisey (2020).

[26] Machiavelli (2005).

realm is in danger, which is why, as the prince thinks, the citizens should not know. Yet, they should know in another sense: the evidence for exposing the lie is available to them. We witness both the practical and epistemic should here.

Machiavelli's prince must be aware of the relevant ironies of the scene. He has a good reason to laugh; his citizens fail, as they should, to see through his plot. Here is the primary source of situational irony: the citizens falsely construe the meaning of the event, which was the plan and purpose of the prince, and hence, they call a crime and conspiracy a blessing. The prince thinks as follows: I killed the police chief, but you do not know the truth, which is just as it should be, a good thing. He has created a situation where bad looks good—a paradigmatic case of irony.[27] The prince may muse: You served me well, I mean by spilling your guts on my piazza.

In this sense, the dramatic characters, the citizens, should and should not know, making this play of ironies on stage so fascinating. The irony of it entails the bivalent use of "should." The characters epistemically should know the truth, which interests the audience (S-effect). But the characters as citizens should not know from the practical political point of view (V-effect). The plot is cruel and the prince is cynical anyway. The Machiavellian example plays with the idea that citizens should not come to know for prudential political reasons. But we can also develop the dramatic plot so that the characters practically, not only epistemically, should come to know.

Example: In Franz Kafka's *The Castle*, K. tries to find his way to the castle on the hill, which he does not know is impossible. He has no relevant evidence. Yet, he practically should know the road because, otherwise, he cannot go there. A critical gap exists between the knowledge possessed by the characters and the audience. K. does not see the ironies of the situation, unlike the audience. K.'s efforts are doomed from the beginning; it is all a wild goose chase. As Kafka develops the drama, K. cannot come to know, and the audience realizes this. The situation is like this:

> K. does not know the road; K. does not know that he cannot know the road; yet he should know it because he wants to go to the castle.
>
> The audience knows that K. cannot come to know the road, and thus he cannot go to the castle.

Such ignorance is crucial from K.'s point of view, yet he misses the irony of his situation. The audience may not miss it, which is the source of dramatic irony. The complex play of "should" and "can" makes the case interesting and worthwhile; without this dialectic, we do not have a claim to think about. The two senses of the expression "should know" provide significance to the scene, and thus motivate us to consider the issue, see its ironies and feel its dramatic weight.

Joanna Garmendia argues that irony is critical.[28] We can express this intuition as follows. Dramatic irony requires that the audience can say the characters *should* act or come to know, which they fail to do. Without this should, we have no dramatic irony. Conventional and other external cases do not satisfy this requirement—they

---

[27] In my (2020a) this idea is central.

[28] See Garmendia (2010). We can also say that irony has its cost and comes with a price; also my (2021a).

do not adequately represent dramatical irony. However, when the audience realizes and says that the characters should know or act, they criticize them. They may pity or sympathize with the characters, but first, they blame. To tell you you *should* react when you do not is criticism. When the audience thinks Gloucester should understand where he is, the audience blames him for needlessly failing. After this, they may pity the man and introduce the soothing S-effect.

References

Aikin, S. (2008). Perelmanian universal audience and the epistemic aspirations of argument. *Philosophy & Rhetoric, 4*, 238–259.

Airaksinen, T. (2017). Nowhere to go Kafka, a philosophical study. *Munich Social Science Review (New Series), 1*, 91–110.

Airaksinen, T. (2020a). Irony and sarcasm in ethical perspective. *Open Philosophy, 3*, 358–368.

Airaksinen, T. (2020b). Mimetic evil: A conceptual and ethical study. *Problemos, 98,* 58–70.

Airaksinen, T. (2021a). Metaphysics, universal irony, and Richard Rorty's "We Ironists". *Humanities, 10*, 1–13.

Airaksinen, T. (2021b). Socrates meets the crocodile, or the ironies of Brecht's and Wuolijoki's Her Puntila und sein Knecht Matti. In T. Airaksinen, M. J. Holler & H. Nurmi (Eds.), *Finnish-German yearbook of political economy, Vol. 3* (pp. 77–96). Verlag Holler.

Airaksinen, T. (2022). Socratic irony and argumentation. *Argumentation, 36*, 85–100.

Bennett, J. (2016). The critical problem of cynical irony *meaning what you say* and ideologies of class and gender. *Social Semiotics, 26*, 250–264.

Brecht, B. (1987). *Mr. Puntila and his man Matti* (J. Willett, Trans.). Methuen Drama.

Burke, K. (1970). *The rhetoric of religion.* University of California Press.

Casali, S. (1995). Tragic irony in Ovid, Heroides 9 and 11. *Classical Quarterly, 45*, 505–511.

Czachesz, I. (2014). *The grotesque body in early Christian discourse*. Routledge.

Dynel, M. (2008). No aggression, only teasing: The pragmatics of teasing and banter. *Lodz Papers in Pragmatics, 4*, 241–261.

Dynel, M. (2016). Comparing and combining covert and overt untruthfulness: On lying, deception, irony and metaphor. *Pragmatics and Cognition, 23*, 174–208.

Euripides. (2017). *Bacchae* (I. Johnston, Trans.). Retrieved from http://johnstoniatexts.x10host.com/euripides/bacchaehtml.html

Friesen, C. J. P. (2014). Dionysus as Jesus: The incongruity of a love feast in Achilles Tatius's Leucippe and Clitophon 2.2. *Harvard Theological Review, 107*, 222–240.

Garmendia, J. (2010). Irony is critical. *Pragmatics and Cognition, 18*, 397–421.

Goldie, P. (2014). *The mess inside: Narrative, emotion, and the mind*. Oxford University Press.

Huson, T. C. (1998). Hegel and the concept of 'tragic irony.' *Southwest Philosophy Review, 14*, 123–130.

Inkpin, A. (2013). Taking Rorty's irony seriously. *Humanities, 2*, 292–312.

Kafka, F. (2019). *The castle (J. R. Underwood, Trans.).* Penguin Modern Classics. Penguin.

Machiavelli, N. (2005). *The prince (Peter Bondanella, Trans.).* Oxford World's Classics. OUP.

Morrisey, C. S. (2020). The truth about Plato's noble lie. Retrieved from https://voegelinview.com/the-truth-about-platos-noble-lie/

Nordquist, R. (2020). Definition and examples of dramatic irony. Retrieved from https://www.thoughtco.com/what-is-dramatic-irony-1690-483

Perelman, C. & Olbrechts-Tyteca, L. (1971). *The new rhetoric*. University of Notre Dame Press.

Phillips, J. (Ed.). (2009). *Tragic play: Irony and theater from Sophocles to Beckett*. Cambridge University Press.

Plato. (2007), *The republic (Desmond Lee, Trans.). Penguin Classic*s. Penguin.
Renegar, V. R., & Goehring, C. E. (2013). A/In (further) defense of irony. *JAC, 33*, 315–324.
Rorty, R. (1989). Private irony and liberal hope. In *Contingency, irony, and solidarity* (pp. 73–95). Cambridge University Press.
Segal, C. (1997). Chorus and community in Euripides' Bacchae. In L. Edmunds & R. W. Wallace (Eds.), *Poet, public, and performance in ancient Greece* (pp. 65–86). Johns Hopkins University Press.
Sophocles. (1919). *Ajax. (R. C. Trevelyan, Trans.).* Retrieved from https://www.classsics.mit.edu/Sophocles/ajax.html
Steinberg, E. R. (1965). K. of the castle: Ostensible land-surveyor. *College English, 27*, 185–189. Retrieved from https://www.jstor.org/stable/373105
Torre, S. (2016). De se thought and communication: An introduction. In M. Garcia-Carpintero & S. Torre (Eds.), *About Oneself: De se thought and communication* (pp. 1–12). Oxford University Press.
Tindale, C. W. (2004). *Rhetorical argumentation: Principles of theory and practice.* Sage.
Williams, G. (1992). Ovid's Canace: Dramatic irony in Heroides 11. *The Classical Quarterly, 42*, 201–209.

Defence is of Much More Importance than Opulence—Adam Smith on the Political Economy of War

Heinz D. Kurz

1 Introduction

There are few people, I surmise, possessed of a better knowledge of Niccolò di Bernardo dei Machiavelli's works, especially his *Il Principe* and his *Discorsi* (see Marchand, 2006 et seq.), than Manfred Holler. He has repeatedly published on Machiavelli's path-breaking contributions to political science (see, inter alia, Holler (2009, 2011)) and has significantly furthered our understanding of them by elaborating on ideas contained therein, using modern analytical tools.

This short paper pays tribute to Manfred, a close friend and esteemed colleague. My focus of attention will, however, not be on Machiavelli, but on Adam Smith. Both authors had the most interesting things to say about war as an option in conflicts among nations or political entities in search of territorial, economic and other forms of power and dominance. The philosopher, historian and social scientist David Hume was full of praise for some of Machiavelli's propositions and wrote that the latter's observations "with regard to the conquest of Alexander the Great … may be regarded as one of those eternal political truths, which no time nor accidents can vary." The reference is to the fact "that such sudden conquests, as those of Alexander, should be possessed so peaceably by his successors, and that the Persians, during all the confusions and civil wars among the Greeks, never made the smallest effort towards the recovery of their former independent government" (Hume, [1777] 1987: 21). The explanation of this surprising fact, Hume surmises, is the kind of government established after the conquest—whether it will follow more "the maxims of the eastern princes" or that of the "western princes". While an eastern prince will "leave no distinction of rank among his subjects, but what proceeds immediately from himself", a western prince, exerting his power "after a milder manner", will instead

H. D. Kurz (✉)
Department of Economics, University of Graz, RESOWI-Zentrum EF, The Graz Schumpeter Centre, 8010 Graz, Austria
e-mail: heinz.kurz@uni-graz.at

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_3

leave "other sources of honour, besides his smile and favour" (ibid: 22). In the former species of government, Hume concludes, "it is impossible ever to shake off the yoke", whereas in the latter, "the least misfortune, or discord among the victors, will encourage the vanquished to take arms" and revolt against the conquerors (ibid: 22).

While Hume called Machiavelli "a great genius", he was no unstinted admirer of his respective studies. These, he objected, were confined to "the furious and tyrannical governments of ancient times, or to the little disorderly principalities of Italy", whereas his observations especially upon monarchical government "have been found extremely defective; and there scarcely is any maxim in his *prince*, which subsequent experience has not entirely refuted" (ibid: 88). Using modern parlance, one might say that Machiavelli has been denied the grace of late birth, or, in Hume's words: he "lived in too early an age of the world, to be a good judge of political truth" (ibid).

Smith, who lived some two and a half centuries later, was lucky in this regard and left to us a multi-layered political economy of war, covering the four stages into which he subdivided the process of socioeconomic civilization up until his time. (This included also remarks on monarchical governments.) Since Smith's political economy of war is not only to be found in Book V of his *Wealth of Nations* (1776), but is spread out over large parts of his oeuvre, one has to pin together the relevant passages in order to get a clear picture of his views on the matter as they developed over time. Hume's objection to Machiavelli that his study is historically contingent, applies, of course, also to Smith. The invention of the atomic bomb and the repeated threat by one of the nuclear powers, the Russian Federation, to use nuclear weapons after Hiroshima and Nagasaki have dramatically changed the situation. As the brutal war waged by the Russian Federation against the Ukraine shows, the danger of a Third World War is real. This no statesman, or student of politics, or taxpayer, can ignore. With a view to World War I that had then just begun, the British journalist Francis Wrigley Hirst, editor of *The Economist*, wrote in his book *The Political Economy of War*, published in 1915: "the fate of civilisation" is at stake (Hirst, 1915: ix). In the nuclear age, in which one side, possessed of a huge arsenal of weapons, has recently repeatedly, and credibly, threatened to use it, what is at stake is no longer the fate of civilization, but that of humankind. This was not so at the time of Adam Smith, which will have to be kept in mind in the following. However, we can nevertheless still learn from the Scotsman today, as this paper argues.

The composition of the essay is the following.[1] Section 2 provides a brief account of the ways in which according to Smith the wealth of a nation (or of groups or single members of it) can be increased, and the role these played in different stages of civilization. Section 3 provides a summary account of the history of warfare, as seen by Smith, in the different stages. The process of civilization is characterized by a self-transformation of society that unleashes forces and tendencies that endanger the continuation of the process from within and renders rich nations susceptible

[1] In this essay I draw freely on a paper in German I wrote on the occasion of the "First Iraq War" (also known as "(Second) Gulf War"), which began with the Iraqi invasion of Kuwait in August 1990; see Kurz (1991); see also Kurz (2022).

to being attacked by "barbarian" nations. Section 4 turns to what Smith calls the "wisdom of the state", without which the process of civilization is doomed to failure. This wisdom Smith then specifies in terms of a number of measures to be taken in order to increase the defence capabilities of a nation. These include, in particular, the instalment of a standing army, which was much debated at Smith's time. Its critics feared that such an army could involve a threat to liberty and freedom. Interestingly, Smith opted in favour of both a standing army and a militia. Section 5 discusses Smith's conviction that wars swiftly tend to corrupt the moral sentiments of people by replacing the "impartial spectator" with a partial one, and the dangers this entails. Section 6 contains some concluding observations.

2 Stages of Civilization and Ways of Accumulating Wealth

In Book V of the *Wealth of Nations* (henceforth *The Wealth* or, simply, WN), Smith defines the duties of the sovereign to consist of the following:

> The first duty of the sovereign, that of protecting the society from the violence and invasion of other independent societies, can be performed only by means of a military force. But the expence both of preparing this military force in times of peace, and of employing it in time of war, is very different in the different states of society, in the different periods of improvement. (WN V.i.a.1)

In the notes taken by a student attending Smith's lectures on jurisprudence in 1762–63, we read: "There are four distinct states which mankind passes thro:—1st, the Age of Hunters; 2dly, the Age of Shepherds; 3dly, the Age of Agriculture; and 4thly, the Age of Commerce" (LJ (A) i.27). This subdivision of the history of human civilisation in four stages recurs in the notes taken on the occasion of the lectures Smith gave in 1766 (see LJ (B) 149) and then in *The Wealth*. In the latter, he provides a brief outline of the history of warfare against the background of the four stages and insists that the process of civilization engenders contradictory tendencies as to the "martial spirit" and military power of a society. The uncertainty about which of these tendencies will prevail in the long run prompts Smith to call on the "wisdom of the state" to muster the necessary defence efforts in order to safeguard society, which necessitates the upkeep and continuous modernization of a sufficiently large and well equipped standing army. Smith follows in this regard a maxim he had put forward in Book IV of *The Wealth*: "defence … is of much more importance than opulence" (WN IV.ii.30). This may be said to be one of the most important maxims in his entire work.

In the Introduction of Book IV of *The Wealth*, "Of Systems of Political Oeconomy", Smith defines the objects of political economy:

> Political economy, considered as a branch of the sciences of a statesman or legislator, proposes two distinct objects; first, to provide a plentiful revenue or subsistence for the people, or more properly to enable them to provide such a revenue or subsistence for themselves; and secondly, to supply the state or commonwealth with a revenue sufficient for the public services. It proposes to enrich both the people and the sovereign. (WN IV.1)

He sees essentially three ways to meet these objectives for a nation as a whole:

- By the conquest of other countries or regions, robbery and tribute payments by subjugated peoples.
- By trade and advantageous exchange according to the mercantilist motto: "buy cheap and sell dear".
- By domestic production via the use of the industry and diligence of people and technical and organizational improvements, and trading parts of the produce with that from other nations.

The first way plays a particularly prominent role in an early phase of socioeconomic development, that is, the first two stages of civilization. It remains important in the mercantile period, in which, however, with the discovery of other continents the second way of enrichment swiftly gains momentum. The third way is the characteristic feature of the age of commerce, in which agriculture and industry are developed, labour productivity is high and rising, and the globalization of the social division of labour is reflected in a growing world market. Provided things go well, Smith is convinced, the process of civilization favours "equality, liberty and justice" (WN IV.ix.3) in society and is accompanied by a sustained improvement of the material provision of people—a rising real income per capita.[2] However, as Smith knows from overwhelming historical evidence, there is no guarantee that the process will go well. There would be no need of a scientific subject called political economy if things were otherwise, that is, if the system, left to its own devices, would consistently generate results that are individually and collectively desirable.[3] The whole purpose of the scientific subject is to elaborate sound economic principles, which, if enacted into law by legislators and implemented by statesmen in terms of economic and social policy, would support the process of civilization and "enrich both the people and the state". The relatively new science of the statesman and legislator has the important task of showing how to contain the forces that endanger the process and strengthen the forces that promote it.

According to Smith, wars waged by less developed nations, possessed of a strong "martial spirit" and envious of the riches of developed nations, which, while possessed of a strong commercial spirit have only a weak martial one, is perhaps the most important external threat to the continuation of the process of civilization.[4] While in the past, wars were mostly fought over natural resources and the fruits of the diligence of a people by nations of similar levels of development, in modern times they concern "barbarian" nations on the one hand and "civilized" ones on the

[2] By "equality" he does not mean equal real incomes, but equal opportunities across all members of society, but he is, of course, aware that this is an ideal that can only be approximated, but not reached.

[3] As will be shown in the sequel, the widespread view that "Adam Smith claimed that nothing more than selfishness is necessary for society to achieve optimal social outcomes" (Schotter, 1985: 2) involves a huge travesty of facts.

[4] There are also forces at work from within society and especially what Smith called the "wretched spirit of monopoly" (WN IV.ii.21), that endanger the system of natural liberty. In this essay they are put on the side; see therefore Kurz (2016).

other. The latter, while materially rich, typically lack sufficient defence capability. As Smith stressed in the *Lectures on Jurisprudence*: "When a country arrives at a certain degree of refinement it becomes less fit for war" (LJ (B) 38). Barbarian countries are, on the contrary, typically relatively poor, but their populations are rich with soldierly virtues. Wealthy nations are therefore in constant fear of assaults from their poorer neighbours. The wealth of a rich nation, Smith stresses, always

> provokes the invasion of all their neighbours. An industrious, and upon that account a wealthy nation, is of all nations the most likely to be attacked; *and unless the state takes some new measures for the public defence, the natural habits of the people render them altogether incapable of defending themselves.* (WN V.i.a.15; emphasis added)

Smith thus attributes a central role to the state for the continuation of the process of civilization: the "wisdom of the state" is needed! In case civilized nations fail to increase sufficiently their military power in order to deter potential invaders, or keep the upper hand in case of military conflict, this process would come to an end and get reversed. The "system of natural liberty" is in danger of becoming the victim of tendencies generated by itself, endogenously. The process of civilization is in jeopardy from within. Smith sees the danger, and while he expresses the hope that it can be contained, he is by no means sure about it. A brief account of his four stages of development and the role of warfare in each of them follows. In it, he draws inter alia on the works of historians from Thucydides and Homer to David Hume. The main reference for the following is Chapter 1 of Book V of *The Wealth.*

3 A Short History of Warfare and the Process of Civilization

In a nation of hunters, Smith insists, the "power of making peace and war … [is] lodged in the whole body of the people" (LJ (B) 26), since "every man is a warrior as well as a hunter" (WN V.i.a.2). While the proportion of people fit for war is large, the actual size of the army is relatively small because of the precarious provision of its members. Hence, while such nations are highly fit for war, they represent no danger for more developed nations. With a view to the situation in the new English colonies, Smith adds: "Nothing can be more contemptible than an Indian war in North America" (WN V.i.a.5). The English and French troops are by far superior to the native Indians and cause harm and bloodshed beyond all measure.

More dangerous for civilized nations are nations of shepherds. They are similarly fit for war as nations of hunters, but they can put up larger armies for basically two reasons: First, "The whole nation, besides being accustomed to a wandering life, even in time of peace, easily takes the field in time of war"; this often includes women, who for example among the Tartars "have been frequently known to engage in battle" (WN V.i.a.3). Secondly, nomads carry with them the main source of their provision, their herds, also in times of war. Compared to hunters, shepherds are possessed of a larger room for manoeuvre and they are also more persevering. Both their economic

activities and the ways of passing their leisure time (running, wrestling, javelin, archery and so on) reflect "the images of war" (WN V.i.a.4). Very much like hunters, they do not need extensive preparation when going to war or a substantial increase in weaponry and sophisticated command structures. State and government play a very modest role and public debt to finance wars is virtually absent.

In nations of farmers in which commerce, handicraft and foreign trade are only poorly developed, people by way of the hardship of their profession are well prepared for the "fatigues of war". However, because of their limited time for leisure, they are less able to train their martial capabilities in a playful way: "They are soldiers, but soldiers not quite so much masters of their exercise" (WN V.i.a.6). There is also the fact that agriculture presupposes a settlement that cannot easily be abandoned by adult men in times of seed and harvest, with the cultivation of the land left entirely to women, the elderly and children. However, military service still does not require much financial support from the government. The military budget of such nations is therefore modest. This was the situation in Ancient Greece, the Roman Empire until the beginnings of the first republic and the European monarchies in feudal times.

Two facts are responsible for a considerable change of affairs in more advanced stages of society. First, with the rise of the manufacturing sector, in which production is continuous, not like in agriculture where it is subject to a rhythm dictated by nature, if men go to war, industrial production as a whole comes to a standstill and so does the stream of income it generates. The implication of this is: "When he [the artisan] takes the field, in defence of the public, as he has no revenue to maintain himself, he must necessarily be maintained by the public" (WN V.i.a.9; see also LJ (A) iv.79).

Secondly, with the rise of manufactures and the money that can be earned and the wealth that can be accumulated in this sector, a new and highly attractive option becomes available to the upper strata of society. While in the past, honour and social reputation could be gained first and foremost by heroic behaviour on the battlefield, and military service, therefore, was one of the noblest duties of the offspring of nobility, now the successful activities of a businessman offers an excellent alternative. Accordingly, Smith opines, "it became inconvenient for the rich to go out to war, from a *principle of avarice* … The merchant who can make 2 or 3000£ at home will not incline to go out to war. But it was an amusement to an ancient knight who had nothing else ado." He goes on: "When the improvement of arts and manufactures was thought an object deserving the attention of the higher ranks, the defence of the state naturally became the province of the lower, because the rich can never be forced to do anything but what they please. … When arts and commerce … begin to be very lucrative, it falls to the meanest to defend the state. This is our present condition in Great Britain" (LJ (B) 335–6; emphasis added). According to Smith this change in the higher ranks' aspirations and lifestyle, and the attempts of the lower ranks to imitate them, is of the utmost importance with regard to the declining capability of civilized nations to defend themselves unless the state takes precautions to prevent this from happening.

The second fact Smith mentions reinforces the first: "the art of war has gradually grown up to be a very intricate and complicated science" (WN V.i.a.10). Military strategy, tactics and weaponry have been developed over time and the average length

of wars has increased. From this follows: "it becomes universally necessary that the publick should maintain those who serve the publick in war, at least while they are employed in that service", since "so very tedious and expensive a service would otherwise be by far too heavy a burden upon them" (WN V.i.a.10). A growing share of public expenses will therefore have to be allocated to national defence.

Smith's view of the impact of the process of civilization on the military fitness of a nation may be summarized in the following way. An increasing social division of labour leads on the one hand to a growing income per capita and an improving material provision of the population. The type of sectors of the economy that gain absolutely and relatively in importance imply, however, that a growing proportion of the people are bound to continuously work in their occupations even in times of war and support the entire nation and the military forces (see WN V.i.a.11; see also LJ (A) iv.79–81 and LJ (B) 37–8). The situation is further aggravated by the fact that in the occupations under consideration military exercises "come to be as much neglected by the inhabitants of the country as by those of the town", with the consequence that "the great body of the people becomes altogether unwarlike" (WN V.i.a.15).

Yet the process of civilization has further detrimental effects on the military spirit and soldierly virtues. It diminishes, Smith opines, "military courage" and, perhaps most importantly, it spreads hedonism: "By having their minds constantly employed on the arts of luxury, [people] grow effeminate and dastardly" (LJ (B) 331).

Smith's deep-seated fear of civilization giving rise to "defects" of the character and leading to effeminacy and dastardliness is huge. The question is: By whom and how can the "disadvantages of a commercial spirit" be contained and the process of civilization salvaged? To Smith, the "wisdom of the state" (WN V.i.a.14) is badly needed; without this wisdom the system of natural liberty would be doomed to failure. Smith's respective reasoning may be subsumed under the heading: *Defence is of much more importance than opulence* (see WN IV.ii.30).

4 On the "Wisdom of the State"

The gradual replacement of the martial by a commercial spirit and the rise to dominance of the *principle of avarice* involves a threat to civilized society that ought to be warded off, but how? What is at stake is a collective good—the integrity and sovereignty of a nation and its inhabitants—and since single individuals are neither willing nor capable of defending this collective good, the community as a whole and its representatives have to step in. Contrary to the civilian sphere in which the division of labour is the result of the working of an "invisible hand" that makes use of the judgement, prudence and self-interest of individuals, in the military sphere, the visible hand of the state is needed. Smith expounds:

> Into other arts the division of labour is naturally introduced by the prudence of individuals, who find that they promote their private interest better by confining themselves to a particular trade, than by exercising a great number. But it is the wisdom of the state only which can render the trade of a soldier a particular trade separate and distinct from all others.

He continues:

> A private citizen who, in time of profound peace, and without any particular encouragement from the publick, should spend the greater part of his time in military exercises, might, no doubt, both improve himself very much in them, and amuse himself very well; but he certainly would not promote his own interest. It is the wisdom of the state only which can render it for his interest to give up the greater part of his time to this peculiar occupation: and states have not always had this wisdom, even when their circumstances had become such, that the preservation of their existence required that they should have it. (WN V.i.a.14)

Smith's above argument alludes already to the first and arguably most important measure to be taken by the state: the establishment of a professional army and its preservation also in times of peace. This is also suggested by the progress in weapons technology due to the deepening of the social division of labour.

> The art of war …, as it is certainly the noblest of all arts, so in the progress of improvement it necessarily becomes one of the most complicated among them. The state of the mechanical, as well as of some other arts, with which it is necessarily connected, determines the degree of perfection to which it is capable of being carried at any particular time. But in order to carry it to this degree of perfection, it is necessary that it should become the sole or principal occupation of a particular class of citizens, and the division of labour is as necessary for the improvement of this, as of every other art. (WN V.i.a.14)

According to Smith, there can be no doubt regarding the "irresistible superiority which a well-regulated standing army has over a militia" (WN V.i.a.28). He is, therefore, convinced that "it is only by means of a standing army, therefore, that the civilization of any country can be perpetuated or even preserved for any considerable time" (WN V.i.a.39). He is also convinced that a standing army is superior to employing mercenaries.

The introduction of a standing army is also suggested by the following fact. Since the offspring of the educated elite is inclined to turn his back to the military service, essentially only members of the lower strata of society become soldiers, who, however, are often unable to read, understand commands and operate sophisticated weaponry. Therefore, the combat efficiency of the army is at risk. A professional army that properly trains its personnel will make a difference. Its instalment implies, of course, a social division of virtues—martial virtues on the one hand and civilian virtues on the other.

However, there is seldom an advantage that does not also carry with it some disadvantage: a standing army poses a potential risk for the polity, as numerous military coups in history show. The risk depends, of course, on a number of factors including the selection mechanism by means of which the people commanding the army are chosen. Institutional arrangements can be installed that minimize the risk. Smith proposes inter alia to complement a standing army with a militia formed by citizens whose main occupations remain their civilian professions. The members of the militia can be expected to care for the protection of the polity including their own private professional interest and thereby deter military encroachments. A militia, Smith adds, has a further advantage: it exposes a large part of the population to regular

military exercises, which keep the martial spirit alive and improve the physical and cognitive strengths of the people.[5]

Smith also stresses that in the course of the development of martial arts, the traditional features of a good soldier change with conventional capabilities such as physical strength gradually losing importance. This is particularly so since the invention of firearms. Now "Regularity, order, and prompt obedience to command, are qualities which, in modern armies, are of more importance towards determining the fate of battles, than the dexterity and skill of the soldiers in the use of their arms" (WN V.i.a.22). The acquisition of these qualities requires troops that are exercised in huge bodies.

Can these plus some other measures guarantee that the threat to the process of civilization can effectively be warded off? Unfortunately, this is not the case. However, "a mere accident" is taken to come to the rescue when the need is greatest. The invention of gunpowder and the "great revolution in the art of war" it caused, Smith is convinced, "is certainly favourable to the permanency and to the extension of the civilization" (WN V.i.a.43–44). He explains this in terms of the cost of such arms, which barbarian nations cannot afford. While this may well be true for some time, with technical progress and the reduction in costs of production, this will not be true forever.

Alas, Smith surprisingly refrains from discussing this possibility. As regards the martial supremacy of civilized nations, he expresses the hope and expectation that they will not use it to invade and conquer other countries, but focus attention on increasing their wealth by means of production and trade. Yet if war is a means to a given end, what would prevent them from using it?

We may at this juncture ask whether Smith's "mere accident" is actually a reification of what he in the *Theory of Moral Sentiments* called "the plan of Providence". There we read about this plan: "The happiness of mankind, as well as of all other rational creatures, seems to have been the original purpose intended by the Author of nature, when he brought them into existence" (TMS III.5.7). And in another passage Smith gives his deism a twist that sounds almost Spencerian: "Nature … seems … to have intended the happiness and perfection of the species" (TMS II.iii.3.2).[6] In case humankind should have reason to rely on "Nature" and its "Author", things would be fine and someone like Putin and his acolytes would be taught a lesson and prevented from making use of their nuclear arsenal. There is something about Smith's respective reasoning that reminds one of whistling in the dark in an attempt to dispel fear. Hopefully, Smith's optimism is justified. As with Wilkins Micawber in Charles Dickens' novel *David Copperfield*, what remains is the hope that "something will turn up".

---

[5] Smith has been attacked in Scotland, where the topic of militia versus standing army gave rise to a heated debate, which involved also several of his friends and opponents (see Kurz, 1991: 116–120). His alleged opposition to a militia and plea for a standing army was criticized as contradicting republican principles. However, as we have just seen, this criticism was unfounded, because he recommended a combination of the two.

[6] Not without some justification, Ronald Coase (1976) therefore called Smith an early evolutionary social scientist.

5 Heroic Characters and the Demise of the "Impartial Spectator"

Smith is convinced that in order to improve the defence capability of a nation the social reputation of the members of the armed forces deserves to be kept high.[7] In the *Theory of Moral Sentiments*, the Stoic ethic he endorsed comes to the fore when he writes:

> The wise and virtuous man is at all times willing that his own private interest should be sacrificed to the public interest of his own particular order or society. He is at all times willing, too, that the interest of this order or society should be sacrificed to the greater interest of the state or sovereignty, of which it is only a subordinate part. He should, therefore, be equally willing that all those inferior interests should be sacrificed to the greater interest of the universe, to the interest of that great society of all sensible and intelligent beings, of which God himself is the immediate administrator and director. (TMS VI.ii.3)

The requested "magnanimous resignation" to the "will of the great Director of the universe" (TMS VI.ii.4) is said to be fully compatible with human nature:

> Good soldiers, who both love and trust their general, frequently march with more gaiety and alacrity to the forlorn station, from which they never expect to return, than they would to one where there was neither difficulty nor danger…. [I]n marching to the former, they feel that they are making *the noblest exertion which it is possible for man to make.* They know that their general would not have ordered them upon this station, had it not been necessary for the safety of the army, for the success of the war. *They cheerfully sacrifice their own little systems to the prosperity of a greater system.* … No conductor of an army can deserve more unlimited trust, more ardent and zealous affection, than the great Conductor of the universe. … A wise man should surely be capable of doing what a good soldier holds himself at all times in readiness to do. (TMS VI.ii.4; emphases added)

Whoever is prepared to defend "liberty and justice, for the sake of humanity and the love of his country", deserves rightly to be admired by his fellow citizens. Smith concludes: "It is this habitual contempt of danger and death which ennobles the profession of the soldier, and bestows upon it, in the natural apprehension of mankind, a rank and dignity superior to that of any other profession" (TMS VI.iii.7).

However, as soon as war breaks out, the institution on which Smith relies most with regard to moral sentiments—the "impartial spectator" in us—gives way to the partial spectator. The result of this is that the "propriety of our moral sentiments is never so apt to be corrupted" (TMS III.3.41): noble sentiments become mean, the love of one's country turns into national prejudice and hatred against other countries, and so on. Smith expounds:

> When two nations are at variance, the citizen of each pays little regard to the sentiments which foreign nations may entertain concerning his conduct. His whole ambition is to obtain the approbation of his fellow-citizens; and as they are all animated by the same hostile passions which animate himself, he can never please them so much as by enraging and offending their enemies. … In war and negotiation, therefore, the laws of justice are very

[7] According to Macfie and Raphael (1976: 18), Smith "seems to have admired heroic characters most".

> seldom observed. Truth and fair dealing are almost totally disregarded. Treaties are violated; and the violation, if some advantage is gained by it, sheds scarce any dishonour upon the violator. (TMS III.3.42; see also LJ (B) 351)

The civilian population is protected least and suffers most: "their lands are laid waste, their houses are burnt, and they themselves, if they presume to make any resistance, are murdered or led into captivity" (TMS III.3.42).

Yet wars do not only have losers, they also have winners, for instance, those who equip the military forces and the producers of provisions whose prices tend to rise. Finally, there are also those who "enjoy, at their ease, the amusement of reading in the newspapers the exploits of their own fleets and armies." These people are commonly dissatisfied with the return of peace, because it puts an end to their amusement, "and to a thousand visionary hopes of conquest and national glory, from a longer continuance of the war" (WN V.iii.37). However, if things go badly for the own army, what was before considered "heroic magnanimity" is then chastised as "extravagant rashness and folly" (TMS VI.iii.28).

In short, wars turn peoples' heads and spoil their character. This can be exemplified in terms of heroes of the war. Because of the enormous amount of recognition and honour bestowed on them, they are inclined to grow their self-admiration immeasurably: "When crowned with success … this presumption has often betrayed them into a vanity that approached almost to insanity and folly" (TMS VI.iii.28). Yet the higher they climb, the harder they fall.

6 Concluding Remarks

In Immanuel Kant's *Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht* (1784) we read: "Aus so krummem Holze, als woraus der Mensch gemacht ist, kann nichts ganz Gerades gezimmert werden" (Out of the crooked timber of humanity, no straight thing can ever be made). Kant, as is well known, was an admirer of Smith. In the latter's reflections referred to here, he turns around and rotates the crooked timber and identifies several sides of it that exist simultaneously and cannot be resolved into one another. Curvature and texture of the always newly sprouting timber remain largely the same as time goes by, but its position and cut do change. What was in the shadow, gets into the light, and what used to be brightly lit disappears into the dark. Formerly trimmed shoots are given free rein, while others are getting pruned. When Smith wrote, the Pre- and Early History of homo sapiens was still unknown. Smith's reflections and speculations about the evolution of human civilization revolve around the rise and fall of social institutions, not about the evolution of the human species. These are taken to be essentially unchangeable. What changes, are the relative importance and the specificity of the various human faculties vis-à-vis altering socioeconomic and environmental circumstances. The selfish and greedy individual that populates large (but by no means all) parts of *The Wealth* (see Book V, in particular), is no other person than the benevolent and devoted member of

the "Commonwealth of all rational beings" of *The Theory of Moral Sentiments*. Smith's political economy of war and peace documents the close connections that exist between his two main works. Without the care and nurturing of the *martial spirit*, the fruits of the *commercial spirit* of modern society are in danger of falling prey to envious neighbours. This danger may be banned or at least contained in an effective and cost-minimizing way by means of a standing army. However, the instalment of such an army does not only involve a division of labour between soldiers and non-soldiers, but also a division of virtues between martial and civilian virtues. This has potentially negative implications for the security of society from within and does not promote the character development of large parts of the male population. Smith, therefore, recommends in addition to a professional army the establishment of a militia that allows the state to achieve several goals at a single stroke. "Defence", Smith insists, "is of much more importance than opulence".

One may wonder whether Vladimir Putin, had he had the privilege of reading Smith's works, would have wished to join "that great society of all sensible and intelligent beings", the "Commonwealth of all rational beings", instead of becoming a criminal and mass murderer from a distance.

References

Coase, R. H. (1976). Adam Smith's view of man. *Journal of Law and Economics, 19*, 529–546.

Hirst, F. W. (1915). *The political economy of war*. J. M. Dent & Sons Ltd.

Holler, M. J. (2009). Niccolò Machiavelli on power. In M. Baurmann & B. Lahno (Eds.), *Perspectives in moral science. Contributions from philosophy, economics, and politics in honour of Hartmut Kliemt* (pp. 335–354). Frankfurt School Verlag. Reprinted in Donskis, L. (Ed.). (2011). *Niccolò Machavelli: History, power, and virtue*. Rodopi.

Holler, M. J. (2011). Machiavelli. In K. Dowding (Ed.), *Encyclopedia of power* (pp. 395–396). Sage Publications.

Hume, D. ([1777] 1987). *Essays moral, political, and literary*. E. F. Miller (Ed.). Liberty Fund.

Kant, I. (1784). Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht. In *Kants Werke* (Vol. VIII). Prussian Academy of Sciences, Berlin 1912.

Kurz, H. D. (1991). Adam Smith über Krieg und Frieden. In *Ökonomie und Gesellschaft*, Jahrbuch 9 (pp. 101–137). Campus.

Kurz, H. D. (2016). Adam Smith on markets, competition and violations of natural liberty. *Cambridge Journal of Economics, 40*, 615–638. (Advanced access publication: 12 March 2015).

Kurz, H. D. (2022, July 4). Adam Smith und der Krieg. In *Frankfurter Allgemeine Zeitung* (No. 152, p. 18).

Macfie, A. L., & Raphael, D. D. (1976). Introduction in Smith (1976a), vol. I of *The Glasgow Edition of the Works and Correspondence of Adam Smith.*

Marchand, J.-J. (Ed.). (2006 et seq.). *Edizione nazionale delle opere di Niccolò Machiavelli* (Vol. 14), Rome.

Schotter, A. (1985). *Free Market Economics.* A Critical Appraisal. New York, St. Martin's Press.

Smith, A. (1976b). *An Inquiry into the Nature and Causes of the Wealth of Nations (WN)*, first published in 1776, vol. II of *The Glasgow Edition of the Works and Correspondence of Adam Smith*, edited by R. H. Campbell, A. S. Skinner and W. B. Todd, Oxford.

Smith, A. (1976a). *The Theory of Moral Sentiments (TMS),* first published in 1759, vol. I of *The Glasgow Edition of the Works and Correspondence of Adam Smith*, ed. by A. L. Macfie und D. D. Raphael, Oxford.

Smith, A. (1977). *Correspondence of Adam Smith (Corr.),* vol. VI of *The Glasgow Edition of the Works and Correspondence of Adam Smith* Smith, A. (1977), edited by E. C. Mossner und I. S. Ross, Oxford.

Smith, A. (1978). *Lectures on Jurisprudence (LJ),* vol. V of *The Glasgow Edition of the Works and Correspondence of Adam Smith*, edited by R. L. Meek, D. D. Raphael and P. G. Stein, Oxford.

Labor Economics

Relative Absence Concerns, Positional Consumption Preferences and Working Hours

Laszlo Goerke

1 Introduction

Individuals care about their own consumption in comparison to that of others. One possible consequence of relative consumption concerns is that the incentives to supply labour are excessive. This is the case since additional income not only enhances consumption possibilities but also improves the own relative position and, thereby, worsens that of others. While the former effect is taken into account by individual decision-makers, the latter impact is ignored. Therefore, relative consumption concerns can justify taxation as a means of internalising this externality.

There also is substantial evidence that higher sickness-related absence by a reference group induces individuals to be absent more. Therefore, absence is associated with what we may call a moral hazard externality, which can be expected to reduce

Manfred Holler's list of publications (in Google Scholar) contains more than 400 entries. Few of them explicitly deal with labour issues—although he has written two textbooks in German on labour economics (Goerke & Holler, 1997; Holler, 1986). My academic interests, which developed during the work as Manfred's chair in Hamburg, attest to the climate of intellectual openness and curiosity, which he created. Braham and Steffen (2008, p. vi) write in their introduction to the *Festschrift* for Manfred's 60th birthday: '(H)e has always made every effort to free his staff from unnecessary administrative burdens and he never burdened anyone with his own work'. This implied that staff members were free to pursue their own research projects, also if only modestly linked to or even without any relationship to Manfred's work. While, to the best of my knowledge, Manfred has not worked on positional concerns, the present analysis was certainly inspired by his attitude of openness and tolerance.

L. Goerke (✉)
IAAEU—Trier University, GLO, IZA, Bonn, CESifo München, Germany
e-mail: goerke@iaaeu.de

Institute for Labour Law and Industrial Relations in the European Union - Trier University, Campus II, D-54286 Trier, Germany

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_4

welfare further. However, upon closer scrutiny, it may also be conjectured that relative absence concerns mitigate or offset a consumption externality. The latter induces individuals to work too much; the former causes them to exert too little effort. Hence, policy prescriptions may be affected by the co-existence of relative absence concerns and positional consumption preferences.

In this chapter we investigate whether relative absence concerns weaken, neutralise or perhaps over-compensate the distortion resulting from relative consumption considerations. If such countervailing effects occur, fostering absenteeism instead of combatting it may be advisable, because absence constitutes a kind of second best instrument. Moreover, if relative absence concerns reduce the distortion resulting from relative consumption considerations, the case for taxation will be weakened or could even become obsolete. Therefore, we furthermore enquire what features characterise optimal income tax rates and optimal sick pay.

In order to analyse these issues, we set up a simple model with an exogenously given number of homogeneous individuals who can determine labour supply and absence. Relative consumption concerns induce individuals to work too much, compared to the Pareto-efficient amount of working hours, assuming the absence level to be given. Endogenising the choice of absence in the next step, while still neglecting absence externalities, to isolate the effects of absence, we can show that individuals still work excessive hours, while there is too little absence from work. The net effect will be positive such that the consumption level remains too high. As a result, allowing for absence in a world with relative consumption concerns does not invalidate the basic prediction resulting from a consumption externality; people work too much. Finally, we extend the model further and assume preferences which are characterised by relative concerns with respect to consumption *and* absence. This additional positional effect gives rise to the moral hazard externality alluded to above. If such absence externalities induce individuals to expand absence, as the available empirical evidence suggests they do, working hours will remain excessive, while absence and consumption can be too low. Therefore, the theoretical analysis predicts that an absence externality will never internalise the impact of relative consumption concerns, in contrast to the conjecture formulated above. The reason is that the utility from absence differs from the utility from leisure such that enhanced incentives to work are never compensated by an augmented inducement to be absent. In consequence, the income tax rate which guarantees a Pareto-efficient outcome is positive, while sick pay is used to internalise the absence externality.

The remainder of the chapter is structured as follows. In Sect. 2, we review related contributions. Section 3 sets out the theoretical model and, in Sect. 4, we compare the market outcome to the Pareto-efficient allocation for alternative specifications of preferences. The main findings are summarised in three Propositions. Section 5 contains some concluding remarks. Most proofs are collected in the Appendix.

2 Related Contributions

The present analysis is primarily related to theoretical and empirical investigations of relative income or consumption effects and empirical contributions considering absence externalities. The subsequent survey focuses on publications in economics.

The first set of relevant studies is motivated by substantial evidence that subjective well-being is influenced by relative income (see Clark et al., 2008; Dolan et al., 2008 for surveys). If utility levels vary with the income or consumption of reference groups, it is but a short way to assume that such externalities also exist with regard to changes in utility. This, in turn, implies that consumption decisions are affected (see, for example, Alpizar et al., 2005; Heffetz, 2011; Kuhn et al., 2011 for evidence). Since expanding consumption generally requires higher income, relative consumption and income concerns tend to distort labour supply (Duesenberry, 1949; Frank, 1985; Schor, 1991). If, in particular, utility declines with the reference level of consumption, labour supply will exceed the Pareto-efficient level (see, inter alia, Persson, 1995; Corneo, 2002; Dupor & Liu, 2003). Additionally allowing for absence, the strength of positional consumption preferences will affect labour supply and absence equally if the latter has the same utility impact as a reduction in working time (cf. Goerke, 2019). This finding is related to the prediction that individuals will no longer necessarily supply excessive amounts of labour in the presence of relative consumption concerns if they also exhibit relative leisure concerns. Because an increase in working time directly affects the relative leisure position, a Pareto-efficient allocation may result, but is unlikely to occur in the presence of both externalities.[1] In many of the contributions which have established the inefficiency of individual choices tax policy has been looked at.[2]

The theoretical prediction that labour supply is excessive in the presence of relative consumption concerns, though not necessarily wealth considerations (Fisher & Hof, 2008), has also found empirical support. Neumark and Postlewaite (1998), for example, employ data from the National Longitudinal Survey of Youth (NLSY) and show that the employment rate of women is higher if either the sister-in-law is employed or the sister's husband has a higher income than the woman's own spouse. Park's (2010) bases his study on the Current Population Survey (CPS) and documents that female labour force participation rises with relative income. Moreover, Pérez-Asenjo (2011) utilises data from the General Social Survey (GSS) and shows that hours of work, the probability of working full-time and labour force participation

[1] See, for example, Seidman (1988), Choudhary and Levine (2006) and Arrow and Dasgupta (2009). Gómez (2008) presents a growth model in which the market equilibrium is efficient if the consumption and leisure externality have the same intensity. In the set-up by Aronsson and Johansson-Stenman (2013), there is asymmetric information with respect to ability, such that the two externalities would not balance out, even if they were equally strong. Alpizar et al. (2005) and Carlsson et al. (2007) present evidence based on hypothetical choice experiments that positional leisure preferences are less pronounced than relative income considerations.

[2] See, inter alia, Duesenberry (1949), Boskin and Sheshinski (1978), Persson (1995), Ireland (1998), Corneo (2002), Gómez (2008), Dodds (2012), Aronsson and Johansson-Stenman (2013), Eckerstorfer (2014), Wendner (2014) and Goerke and Neugart (2021).

decline with relative income. Bracha et al. (2015) conducted a laboratory experiment and informed a subset of the participants that they earn half of the amount paid to other participants performing the same tasks. They present evidence that information about relative pay tends to reduce the labour supply of those male subjects paid a lower wage unless a strong justification for unequal pay was provided. Breza et al. (2018) used a field experiment to, inter alia, analyse the effects of relative wages on attendance. Workers who receive a relatively low wage are more likely not to turn up for work, whereas there are no such effects observable for workers who earn high relative wages. In sum, the theoretical notion that relative consumption or income concerns can cause excessive labour supply is well established and there is empirical evidence corroborating this prediction.[3]

The second relevant strand of literature relates to empirical studies of relative absence effects.[4] Ichino and Maggi (2000) employ firm-level data from a large Italian bank. Using changes between branches and different parts of the country to identify social interactions, they find a significant positive impact of the average number of absence periods on individual absence behaviour. De Paola (2010) also utilises firm-level data for a much smaller sample of Italian public sector employees. Relying on an instrumental variable approach, she identifies positive spill-over effects of the absence rates of co-workers.

In a further important paper, Lindbeck et al. (2016) analyse all absence periods lasting longer than 14 days in Sweden for the period 1996–2002. For a variety of identification strategies, they show that the average absence duration in a narrowly defined neighbourhood significantly increases the absence of the individual under consideration. Also employing data from Sweden, Hesselius et al., (2009, 2013) and Johansson et al. (2019) base their studies on a social experiment that took place in Göteborg at the end of the 1980s. At that time, employees could claim sick pay for a week without presenting a medical certificate. This requirement was relaxed for a randomly chosen subgroup of employees. Hesselius et al. (2009) find, inter alia, that the proportion of individuals affected in an employee's workplace had a positive and significant impact on the duration of absence. Johansson et al. (2019) additionally

[3] There are further contributions which point into the same direction. Aronsson et al. (1999) analyse the implications of interdependent labour supply behaviour for estimated labour supply elasticities, using repeated cross-sectional data from Sweden. They find that average working hours in a reference group raise individual labour supply. Pingle and Mitchell (2002) set up a hypothetical choice experiment. They present individuals with combinations of working time and income and report that the average level of hours worked affect individuals' choices.

[4] Palme and Persson (2020, Sect. 4) concisely review pertinent empirical studies. Miraglia and Johns (2021) provide a much broader survey of the literature on social determinants of absence behaviour, also including contributions from economics. To the best of our knowledge, the implications of relative absence concerns have not yet been analysed in a theoretical model. Somewhat related to our analysis, Skåtun and Skåtun (2004) analyse an efficiency wage model in which individuals can choose hours of work. The authors interpret this choice as a decision about absence behaviour. They assume that fewer hours worked by colleagues raise the workload of individuals and, hence, reduce the individual's working hours as well. The main prediction of the model is that, in contrast to traditional shirking frameworks, employment may be higher in the presence of efficiency wages than in their absence.

demonstrate that these externalities vary with the type of job. Hesselius et al. (2013), furthermore, look at employees living in bordering municipalities. They also find substantial evidence of positive absence spill-overs.

Moving from Sweden to Norway, Dale-Olsen et al. (2015) use a variation in marginal income tax rates and, hence, in net wages in 2006 to identify a change in the incentives to attend work. In a sample of male employees, they observe substantial positive effects of the colleagues' average absence on an individual's own duration of absence. Further evidence of peer effects in Norway is provided by Godøy and Dale-Olsen (2018). They show that an arguably exogenous change in the leniency of certifying sick leave alters an employee's absence behaviour and also that of colleagues who are unaffected by the variation in attesting an illness in the same direction. A final piece of evidence is provided by Bradley et al. (2007) who investigate the behaviour of school teachers in Queensland, Australia. They identify positive interaction effects of illness-related absence by focusing on individuals who move between schools.[5]

In sum, there is consistent evidence originating from various countries and approaches that the absence level of people who are employed in the same workplace has a positive impact on the absence of the individuals under consideration. Such externalities are sometimes interpreted as shirking (e.g. Bradley et al., 2007; Ichino & Maggi, 2000) or as resulting from fairness or reciprocal concerns (i.e. by Dale-Olsen et al., 2015; Hesselius et al., 2013), while there is no evidence that they arise due to the spread of contagious diseases.

Based on the above contributions, we subsequently assume preferences, which ensure that (1) an individual's labour supply rises with reference consumption, and (2) the absence level of an individual increases with the absence of their reference group.

3 Model

In this section, we initially outline the foundations of our analysis, subsequently delineate the details of the model, then derive the market outcome and, finally, describe the Pareto-efficient allocation.

Foundations

We consider a single-period setting, with a given number of identical individuals who decide about working time. Therefore, adjustments in labour supply only take place at the intensive margin. Moreover, intertemporal repercussions of positional consumption preferences and relative absence concerns are ruled out. There is full

[5] Bradley et al. (2014) investigate the impact of a move from temporary to permanent employment on absenteeism for public sector employees in Australia. In some of their specifications they include an indicator of the average absence level at the employee's workplace. The estimated coefficients are consistently positive and significant.

employment and the only actors are workers and firms. The latter cannot influence the output price and are profitable at the wage prevailing in the absence of positional or relative preferences. This assumption ensures that firms can pay wages above the wage resulting in a world without comparison effects. To close the model and retain the homogeneity assumption, profits are redistributed equally to all workers. Since this profit component of income is exogenous from each individual's perspective, the profit income does not affect the impact of relative concerns.

The above assumptions, and further ones outlined below, ensure that labour supply will be excessive in the presence of positional consumption preferences if there are no relative absence concerns (see Sect. 4.1). Therefore, the set-up allows us to isolate the impact of relative absence concerns and their interaction with positional consumption preferences.

Relaxing one or more of the assumptions could, for example, imply that labour supply in the absence of relative absence concerns is no longer excessive. To illustrate, suppose that the labour market is not perfectly competitive, but that market power either by employees or firms reduces the employment to below the level prevailing in a setting without such distortion. In such cases, positional consumption preferences can bring the economy closer to efficiency or even guarantee an efficient outcome (Goerke & Hillesheim, 2013; Goerke & Neugart, 2021). Relaxing the assumption of homogeneous individuals would imply that there may be many Pareto-efficient allocations. Therefore, the effect of relative absence concerns could crucially depend on the benchmark, which is selected to evaluate the market outcome. Finally, relaxing, for example, the assumption of a given number of individuals could result in excessive labour supply not only at the intensive but also, or instead, at the extensive margin, depending on, inter alia, the firms' production technology.

In consequence, all of the simplifying assumptions laid out above may determine the findings presented in Sect. 4, and the policy conclusions derived from them. However, the simplifications greatly help to isolate and understand the basic mechanisms governing the interaction between positional consumption preferences and relative absence concerns.

Set-up

The large number of identical individuals can divide up their time endowment, which we set to unity, into actual working time, $h - a$, absence, a, and leisure, $1- h$. Actual working time, also referred to as *effort*, is the difference between labour supply or contractual working hours, h, and absence, a. Individuals derive utility from consumption of the single commodity, c, leisure, $1- h$, and absence, a. Moreover, utility depends on the choices of a reference group, namely the average levels of consumption and absence, $\overline{c}$ and $\overline{a}$. Overall utility, Z, is then specified as

$$Z(h, a) = u(c, \overline{c}) - H(h) + v(a, \gamma \overline{a}) \tag{1}$$

The separability between the (sub-) utility from consumption, u, the disutility, H, from contractual working hours, h, and the (sub-) utility, v, from absence, a, substantially simplifies subsequent computations, without imposing too much structure on preferences.

Utility, u, is increasing in personal consumption, c, at a decreasing rate ($u_1 > 0 > u_{11}$), where subscripts denote partial derivatives. In line with our motivation, individuals are characterised by envy with regard to income (Dupor & Liu, 2003). Accordingly, utility decreases with the consumption level of the reference group, $\overline{c}$ ($u_2 < 0$). This ensures that working hours chosen individually are excessive, as demonstrated by Cahuc and Postel-Vinay (2005), Persson (1995), Alvarez-Cuadrado (2007), Pérez-Asenjo (2011), Dodds (2012) and Goerke (2019), inter alia. Moreover, utility declines with (contractual) working hours at an increasing rate (H', $H'' > 0$), reflecting the positive but decreasing marginal utility from leisure. Finally, absence raises utility, albeit at a decreasing rate ($v_1 > 0 > v_{11}$). Such positive absence effects can arise, for example, because people who are ill gain extra utility from not having to work. Alternatively, absence can be viewed as being unrelated to health and to be due to shirking. In the present setting, there is no need to precisely determine the cause of absence, because our findings rely on the assumption that the utility from absence, a, differs from the utility due to leisure, $1 - h$. This will certainly be true in the two polar cases outlined above, particularly if shirking involves, for example, feelings of guilt or restricts the range of activities which can be undertaken while being officially ill.[6]

In Eq. (1), utility from absence, v, also depends on the reference group's absence level, $\overline{a}$. The non-negative parameter γ measures the strength of this absence externality. The empirical evidence summarised above suggests that higher absence by a reference group increases an individual's absence level, implying that $\partial v_1/\partial \overline{a} > 0$ holds. This effect may arise because the reference level defines a social norm or focal point. Alternatively, our specification may capture the impact of additional workload arising if colleagues are absent. This extra effort will raise the disutility from work, such that the gain due to own absence rises. Irrespective of the mechanism underlying the absence externality, the empirical evidence does not provide consistent information concerning the direction of the direct utility effect of $\overline{a}$, that is, the sign of $\partial v/\partial \overline{a} = \gamma\, \partial v/\partial(\gamma \overline{a}) = \gamma v_2$.[7]

[6] Absence can also have detrimental effects on future wages and employment (see, e.g., Hansen, 2000; Hesselius, 2007; Markussen, 2012; Scoppa & Vuri, 2014). While we do not model such consequences explicitly, one feasible short-cut in order to incorporate them into the model is the above assumption that utility from absence is distinct from that due to leisure.

[7] Carrieri (2012) found that a higher sickness level of a reference group reduces well-being. If (1) higher sickness induces people to be absent more and (2) utility from absence can be approximated by subjective well-being, Carrieri's (2012) result suggests $v_2 < 0$. However, this line of argument may be problematic, given survey results that positional concerns with regard to health are relatively weak (cf. Solnick & Hemenway, 2005; Grolleau & Saïd, 2008; Wouters et al., 2015). These findings from surveys contrast with evidence from panel data for Australia (cf. Mujcic & Frijters, 2015) according to which the self-assessed health status of a peer group is consistently and strongly associated with a reduction in life satisfaction.

Empirically, a substantial fraction of employees is observed never to be absent from work (Frick & Malo, 2008). In order to ensure interior choices of working hours and absence, we postulate that $u_1(c \to 0) \to \infty$, $H'(h \to 0) \to 0$, $H'(h \to 1) \to \infty$ and $v_1(a \to 0) \to \infty$ hold. Further, marginal utility from consumption, u_1, and absence, v_1, decline with a general rise in consumption, respectively, absence. This requires $u_{11} + u_{12} < 0$ and $v_{11} + \gamma v_{12} < 0$ and guarantees stability of the equilibrium, together with the restrictions on H and the production function (see below). Following, for example, Dupor and Liu (2003), we additionally assume that a general rise in consumption makes the individual under consideration better off ($u_1 + u_2 > 0$ for $\mathrm{d}c = \mathrm{d}\overline{c}$) and decreases the marginal utility from consumption ($\mathrm{d}(u_1 + u_2)/\mathrm{d}c = u_{11} + 2u_{12} + u_{22} < 0$). Similar restrictions with respect to utility from absence are imposed, implying that $v_1 + \gamma v_2 > 0$ and $\mathrm{d}(v_1 + \gamma v_2)/\mathrm{d}a = v_{11} + 2\gamma v_{12} + \gamma^2 v_{22} < 0$ hold for $\mathrm{d}a = \mathrm{d}\overline{a}$. Since the number of individuals is fixed, we can finally, and without loss of generality, normalise their number to unity.

Production takes place in a representative firm which produces the single consumption good with labour as the sole factor. Output and consumption are given by $c = f(h - a)$, where f constitutes the production function which is increasing in effort, $h - a$, at a decreasing rate ($f' > 0 > f''$). If effort is zero, $h - a = 0$, so will be output ($f(0) = 0$), while the first unit of effort will be infinitely productive ($f'((h - a) \to 0) \to \infty$). Because the output price is constant we also normalise it and set it equal to unity to save on notation.

Market Equilibrium

In market equilibrium, all individuals are employed and earn a wage, w, per unit of working time. Moreover, they may receive sick pay, s, per time unit of absence, as is the case in most OECD countries (OECD, 2010, pp. 128 f). While wage income is generally taxed, the picture relating to sick pay is more mixed (see MISSOC (2021) for European Union and EFTA countries). Therefore, we assume that a linear income tax is levied on wages at the rate t, $0 \leq t < 1$, while sick pay remains untaxed and does not exceed the net wage, $0 \leq s \leq w[1 - t]$.[8] Tax receipts are returned to individuals in a lump-sum manner. The respective payment T equals $w[h - a]t$ in equilibrium. Furthermore, to close the model, individuals obtain profit income, π.

Since individuals cannot save, consumption, c, and total net income coincide and are given by[9]

$$c = w[1-t][h-a] + sa + T + \pi \tag{2}$$

[8] The subsequent findings are unaffected by the assumption that sick pay is untaxed, unless noted below (cf. Proposition 3). To focus on relative absence and consumption concerns, the model developed below is static. As mentioned above, there is substantial evidence that sickness-related absence has detrimental long-term labour market effects (Hansen, 2000; Hesselius, 2007; Markussen, 2012; Scoppa & Vuri, 2014). An alternative or additional way of including this empirical observation in the present static setting is the assumption that sick pay is less than the net wage, i.e. $s < w[1 - t]$.

[9] Note that terms in square brackets describe multiplicative components, while parentheses indicate a functional dependence.

Maximisation of utility $Z(h, a)$ with respect to h and a, subject to the individual budget constraint (2), taking as given average consumption, $\overline{c}$, average absence, $\overline{a}$, the wage, w, lump-sum returns from tax authorities, T, and profits, π, yields two first-order conditions (for the second-order conditions, see Appendix 1).

$$Z_{\mathrm{h}} = u_1(c, \overline{c})w[1 - t] - H'(h) = 0 \tag{3a}$$

$$Z_{\mathrm{a}} = u_1(c, \overline{c})[s - w[1 - t]] + v_1(a, \gamma\overline{a}) = 0 \tag{3b}$$

The individual choice of working hours, denoted by h^{m} in market equilibrium, results from the trade-off between the additional utility due to higher net income and the loss in utility from less leisure. The duration of absence a^{m} in equilibrium balances the utility change resulting from the loss in income with the utility gain from not having to work.

The representative firm covers the costs of sick pay, as it is the case for shorter absence spells in many OECD countries (cf. OECD, 2010, pp. 128 f; MISSOC, 2021). Hence, profits are given by

$$\pi(h) = f(h - a) - w[h - a] - sa \tag{4}$$

Maximisation of profits, $\pi(h)$, with respect to contractual working hours, h, yields

$$\pi_{\mathrm{h}} = f'(h - a) - w = 0 \tag{5}$$

The second-order condition for a maximum is guaranteed by the strict concavity of the production function ($\pi_{\mathrm{hh}} = f'' = -\pi_{\mathrm{ha}} < 0$).

In equilibrium, wage adjustments ensure that labour demand and labour supply coincide. Accordingly, working hours, h, absence, a, the reference values of consumption, $\overline{c}$, and of absence, $\overline{a}$, as well as the wage, w, are endogenous variables. Moreover, tax payments equal the lump-sum transfer, T, in equilibrium and consumption is affected by changes in endogenous variables via the resulting variation in profits. In Appendix 2, we show that given the restrictions on the utility function imposed above, the Jacobian determinant, $|J|$, of the system of Eqs. (3a), (3b) and (5), taking into account the above repercussions, is negative.

Given a balanced-budget requirement and taking into account profit income (cf. Eq. (17) in the Appendix), the derivatives of the first-order conditions (3a) and (3b) with respect to the tax rate, t, and sick pay, s, are given by $Z_{\mathrm{hs}} = 0$, $Z_{\mathrm{at}} = -Z_{\mathrm{ht}} = u_1(c, \overline{c})w$ and $Z_{\mathrm{as}} = u_1(c, \overline{c})$. This implies that working hours and absence rise with sick pay ($\mathrm{d}h^{\mathrm{m}}/\mathrm{d}s$, $\mathrm{d}a^{\mathrm{m}}/\mathrm{d}s > 0$; see also Appendix 6), absence increases with the tax rate ($\mathrm{d}a^{\mathrm{m}}/\mathrm{d}t > 0$), while the effect of a tax rate change on working hours is determined by the sign of $v_{11}(a, \gamma\overline{a}) - u_{11}(c, \overline{c})sf'(h - a)$ and, hence, ambiguous for $s > 0$.

Pareto-Efficiency

In a Pareto-efficient allocation, all individuals are treated identically and there will not be sick pay because consumption can be determined directly. Therefore, Pareto-efficiency is characterised by a maximum of utility Z, subject to the output constraint, $c = f(h - a)$, and the restriction that personal consumption and absence levels, c and a, as well as working hours, h, coincide with their respective averages, $\overline{c}$, $\overline{a}$, and $\overline{h}$. This implies that $\mathrm{d}c = \mathrm{d}\overline{c}$, $\mathrm{d}a = \mathrm{d}\overline{a}$ and $\mathrm{d}h = \mathrm{d}\overline{h}$ hold. From the output constraint, the relationship between consumption and working hours can be derived.

$$\frac{\mathrm{d}c}{\mathrm{d}h} = -\frac{\mathrm{d}c}{\mathrm{d}a} = f'(h - a) > 0 \tag{6}$$

Differentiation of

$$\Gamma(h, a) = u\big(c(h, a), \overline{c}\big(\overline{h}, \overline{a}\big)\big) - H(h) + v(a, \gamma\overline{a}) \tag{7}$$

with regard to hours, h, and absence, a, yields (see Appendix 3 for the second-order conditions):

$$\Gamma_{\mathrm{h}} = [u_1(c, \overline{c}) + u_2(c, \overline{c})]f'(h - a) - H'(h) = 0 \tag{8a}$$

$$\Gamma_{\mathrm{a}} = -[u_1(c, \overline{c}) + u_2(c, \overline{c})]f'(h - a) + v_1(a, \gamma\overline{a}) + \gamma v_2(a, \gamma\overline{a}) = 0 \tag{8b}$$

The Pareto-efficient number of working hours, h^*, results from the trade-off between the additional utility from the higher output and, hence, consumption on the one hand, and the utility reduction due to less leisure on the other (cf. Eq. (8a)). The utility gain from greater consumption, in turn, consists of a direct, positive effect and an indirect, negative one, because higher consumption by other individuals reduces utility due to relative consumption concerns ($u_2 < 0$), ceteris paribus. The Pareto-efficient absence level, a^*, arises from a similar trade-off as it applies to working hours (cf. Eq. (8b)). Since all individuals are treated identically, the Pareto-efficient allocation is uniquely defined.[10]

The Pareto-efficient allocation can be attained in a market economy in which working hours and absence cannot be determined directly by setting the tax rate, t, and sick pay, s, in such a manner that the objective $\Gamma(h, a)$ (cf. Eq. (7)) is maximised. This is feasible because taxes and sick pay affect individual choices (in accordance with Eqs. (3a) and (3b)).

[10] It could be argued that absence has a distinct, positive utility effect. If, therefore, preferences were given by $u(c, \overline{c}) - H(h, a) + v(a, \gamma\overline{a})$, where the partial derivatives are $\tilde{H}_1 < 0 < \tilde{H}_2$, the nature of the first-order conditions for individual choices and for the characterisation of Pareto-efficiency would not be altered under mild additional restrictions (see Appendix 4). Hence, the findings derived below are unlikely to be affected.

4 Comparisons of Outcomes

4.1 Working Hours as Sole Choice Variable

In the first step, we assume that individuals can decide only about working hours, so that there is no absence ($a = v_1 = v_2 = 0$). The market equilibrium is then defined by the combination of Eqs. (5) and (3a) for $a = s = 0$ and Pareto-efficiency by Eq. (8a), which we rewrite as (9b) for ease of comparison, while imposing $v_1 = v_2 = 0$.

$$f'(h^{\mathrm{m}}) = \frac{H'(h^{\mathrm{m}})}{u_1(c^{\mathrm{m}}, \bar{c})[1-t]} \tag{9a}$$

$$f'(h^*) = \frac{H'(h^*)}{u_1(c^*, c^*) + u_2(c^*, c^*)} \tag{9b}$$

Inspection of (9a) and (9b) shows that:

Proposition 1

Assume that individuals only decide about working hours in market equilibrium.

(a) If income is untaxed, working hours will be excessive.
(b) The tax rate, $t(a = 0)$, which ensures that individuals choose the optimal number of working hours, is given by

$$0 < t(a = 0) = -\frac{u_2(c^*, c^*)}{u_1(c^*, c^*)} < 1. \tag{10}$$

Proof Part (a) follows from the comparison of (9a) and (9b) for $u_2 < 0$, from the strict concavity of f in h, the assumption that u_1 and $u_1 + u_2$ decrease in consumption c, and the production constraint (cf. Eq. (17) in Appendix 2). Substituting $t(a = 0)$ into (9a) shows that this equality will then hold for working hours h^* which are implicitly defined by (9b). As, moreover, tax payments $w[h - a]t$ equal the lump-sum transfer, T, and profits are returned to individuals, their income and, hence, consumption levels will be the same as in the Pareto-efficient allocation, given $h^{\mathrm{m}} = h^*$ and $c^{\mathrm{m}} = c^*$. Since $u_1 + u_2 > 0 > u_2$ for any given combination of h and a, the tax rate $t(a = 0)$ is less than unity. ■

Proposition 1 indicates that individuals work too many hours, h, because they do not take into account that an increase in h decreases other individuals' relative consumption position. This prediction of excessive labour supply is well established and variants of it have been derived, for example, by Seidman (1988), Persson (1995), Ljungqvist and Uhlig (2000), Corneo (2002), Dupor and Liu (2003), Cahuc and Postel-Vinay (2005), Alvarez-Cuadrado (2007), Pérez-Asenjo (2011) and Goerke and Hillesheim (2013). Since labour demand is unaffected by relative consumption concerns, excessive supply of working hours translates into too much equilibrium

effort. Because consumption rises with effort, the same is true for the consumption level. Turning to the optimal tax rate, it can be noted that the gain to society from a general rise in consumption is given by $u_1 + u_2$, because consumption of the reference group also rises. If an individual decides about consumption, the marginal gain is only u_1 because the variation in the consumption of other individuals' reference groups is ignored. Therefore, individual consumption—or hours—decisions neglect a fraction u_2/u_1 of the utility change. As established in a variety of analyses (e.g. Alvarez-Cuadrado, 2007; Aronsson & Johansson-Stenman, 2010, 2013, 2018; Dupor & Liu, 2003; Liu & Turnovsky, 2005; Ljungqvist & Uhlig, 2000; Persson, 1995), the tax rate on income which mimics this fraction, induces an individual to choose the optimal number of working hours.

4.2 Absence of Absence Externality

In this sub-section, we assume that individuals decide about working hours and the duration of absence, but we continue to disregard absence externalities ($\gamma = 0$). To isolate the impact of envy, we initially set sick pay equal to zero ($s = 0$) and consider its effects at the end of this sub-section. Imposing $\gamma = 0$ in the conditions characterising the market equilibrium (3a), (3b) and the Pareto-efficient outcome (8a), (8b), and combining (3a) and (3b) with (5) yields

$$u_1(c^{\mathrm{m}}, \overline{c})\, f'(h^{\mathrm{m}} - a^{\mathrm{m}})[1 - t] - H'(h^{\mathrm{m}}) = 0 \tag{11a}$$

$$-u_1(c^{\mathrm{m}}, \overline{c})\, f'(h^{\mathrm{m}} - a^{\mathrm{m}})[1 - t] + v_1(a^{\mathrm{m}}) = 0 \tag{11b}$$

$$\left[u_1\left(c^*, c^*\right) + u_2\left(c^*, c^*\right)\right] f'\left(h^* - a^*\right) - H'\left(h^*\right) = 0 \tag{12a}$$

$$-\left[u_1\left(c^*, c^*\right) + u_2\left(c^*, c^*\right)\right] f'\left(h^* - a^*\right) + v_1\left(a^*\right) = 0 \tag{12b}$$

From the comparison of these equations, we obtain

Proposition 2

Assume that individuals decide about working hours and absence in market equilibrium, sick pay is zero ($s = 0$), and there is no absence externality ($\gamma = 0$).

(a) If income is untaxed, working hours, effort and consumption will be excessive, while there will be too little absence from work.
(b) The tax rate $\hat{t} := t(s = \gamma = 0)$, which ensures that individuals choose the optimal number of working hours and the optimal duration of absence, is given by

$$0 < \hat{t} = -\frac{u_2(c^*, c^*)}{u_1(c^*, c^*)} < 1. \tag{13}$$

Proof: See Appendix 5.

In order to provide intuition for Proposition 2, note that individuals will be absent from work even if they do not obtain income while being away ($s = 0$) because absence increases utility directly. Relative to the Pareto-efficient outcome, however, absence is too low. This is the case since absence reduces income and, thus, consumption possibilities. When trading off the gain from being absent, i.e. the increase in 'absence utility', v, with the costs in the form of lower consumption, an individual will not take into account that lower consumption makes all other individuals better off because of the existence of envy, as captured by the term u_2 (< 0) in (12b), which is not contained in (11b), describing individual behaviour.

The combination of Eqs. (11a), (11b) and (12a), (12b) shows that the relationship between working hours and absence both in market equilibrium and in the Pareto-efficient allocation is governed by the equality of the marginal disutility of work, $-H'(h)$, and the marginal utility from absence, $v_1(a)$. As argued above, absence is too low in market equilibrium. Hence, the marginal utility from absence is higher than in a Pareto-efficient allocation. Consequently, also the marginal disutility from extra hours must be higher, that is, less leisure must be consumed in market equilibrium relative to the efficient allocation. Therefore, working hours in market equilibrium are excessive, whereas absence is too low, and effort and consumption must surely be too high. Proposition 2, furthermore, states that a single tax instrument suffices to ensure the efficiency of two endogenous variables, namely working hours and absence. This is the case because the choice of absence is not distorted, for a given number of working hours.

Moving beyond the narrow confines of our theoretical set-up, the above prediction of insufficient absence could also be interpreted in light of the debate about presenteeism. In order to do so, the model would have to be extended to allow for truly sickness-related absence. Assume, therefore, that each individual is sick for some time and then has to decide whether to attend work or to be absent. If the gain from absence is highest when sick, too little absence implies that individuals will sometimes be present at work although they are ill. Hence, the existence of relative consumption concerns would imply that people not only supply too many hours but also go to work too often when ill. Consequently, relative consumption concerns in the form of envy can be argued to cause presenteeism.

Allowing for an Exogenously Given Level of Sick Pay

Subsequently, we relax the restriction that there is no sick pay. If one instrument, the income tax rate $\hat{t}$, suffices to induce individuals to make efficient choices in market equilibrium (cf. Proposition 2, Part (b)), a second market intervention can only cause inefficiencies. Therefore, any exogenously given, non-zero sick pay, s, will cause a distortion. The reason is that sick pay only affects the absence decision, but does not alter the choice of working hours (cf. Eqs. (3a) and (3b)). In consequence, irrespective

of the tax rate, the market equilibrium cannot be efficient in the presence of sick pay if there is no absence externality.

If sick pay is positive, contractual working hours and absence will be higher in market equilibrium than for $s = 0$, for a given tax rate, while effort will be lower (see Appendix 6).[11] Sick pay has no income effect in equilibrium because it is paid for by firms. Hence, the rise in labour income is compensated by the fall in profits, such that the overall impact on an individual's income and consumption is zero. However, higher sick pay raises the marginal gain from absence. Thus, the substitution effect implies that absence rises, such that production and consumption will decline. To mitigate this detrimental impact on utility, contractual hours are increased. Their rise will be less pronounced than the increase in absence because of the disutility of contractual hours, $H(h)$. Consequently, work effort declines.

Absence in market equilibrium without sick pay is too low because individuals ignore the impact of absence on the reference income. The absence level in a market equilibrium in which sick pay is positive will be higher ($a^{\mathrm{m}}(s > 0) > a^{\mathrm{m}}(s = 0)$). The Pareto-efficient allocation, however, is characterised by the absence of sick pay. Therefore, absence resulting in market equilibrium with sick pay may be higher or lower than the Pareto-efficient amount. Furthermore, working hours will be excessive in market equilibrium without sick pay, s, and will rise with s. Since their Pareto-efficient number is independent of sick pay, working hours will surely be excessive in the presence of sick pay. Finally, the effort level resulting in market equilibrium with sick pay may be higher or lower than the Pareto-efficient amount since (1) effort is excessive if sick pay is zero and (2) effort declines with sick pay.

4.3 Simultaneous Existence of Consumption and Absence Externalities

Propositions 1 and 2 show that contractual working hours and effort are excessive in a world with envy. We next scrutinise whether the prediction continues to apply if there also is an absence externality. To do so, we initially consider the impact of a greater reference level of absence, $\overline{a}$, on individual choices, holding constant the wage, that is, for a given market outcome. The impact of $\overline{a}$ on individual choices is determined by $Z_{h\bar{a}} = 0$ and $Z_{a\bar{a}} = \gamma v_{12}$. Using $Z_{ha} > 0$ and $Z_{hh}Z_{aa} - (Z_{ha})^2 > 0$ (cf. Appendix 1), we obtain

$$\frac{\mathrm{d}a^{\mathrm{m}}}{\mathrm{d}\overline{a}}_{|\mathrm{d}w=0} = -\gamma v_{12}\frac{Z_{hh}}{Z_{hh}Z_{hh} - (Z_{ha})^2} \tag{14a}$$

$$\frac{\mathrm{d}h^{\mathrm{m}}}{\mathrm{d}\overline{a}}_{|\mathrm{d}w=0} = \gamma v_{12}\frac{Z_{ha}}{Z_{hh}Z_{hh} - (Z_{ha})^2} \tag{14b}$$

[11] Since effort is too low in market equilibrium and declines with sick pay, such payments can be argued to reduce presenteeism (see Pichler and Ziebarth (2017) for according empirical evidence).

$$\frac{\mathrm{d}(h^{\mathrm{m}} - a^{\mathrm{m}})}{\mathrm{d}\overline{a}}_{|\mathrm{d}w=0} = \gamma v_{12} \frac{Z_{\mathrm{ha}} + Z_{\mathrm{hh}}}{Z_{\mathrm{hh}} Z_{\mathrm{hh}} - (Z_{\mathrm{ha}})^2} = \gamma v_{12} \frac{u_{11} w[1-t]s - H''}{Z_{\mathrm{hh}} Z_{\mathrm{hh}} - (Z_{\mathrm{ha}})^2} \tag{14c}$$

If we take as our starting point the overwhelming empirical evidence that a higher absence level of a reference group raises absence by an individual (see Sect. 2), the derivative in (14a) and, consequently v_{12}, are positive. Therefore, working hours in market equilibrium also increase with the absence of the reference group, while effort declines.

Assuming $v_{12} > 0$, we can next compare the market outcome and the Pareto-efficient allocation and analyse the importance of relative absence concerns. Additionally, Proposition 3 characterises the tax rate and level of sick pay that induce Pareto-efficient choices.

Proposition 3

Assume that individuals decide about working hours and absence and there is an absence externality ($\gamma > 0$).

(a) If the tax rate is zero ($t = 0$), sick pay is non-negative ($s \geq 0$), and higher absence by the reference group does not raise utility from absence ($v_2 \leq 0$), working hours in market equilibrium will be excessive, while the differences between Pareto-efficient and market outcomes with respect to absence and consumption are indeterminate.
(b) If a higher absence level by the reference group increases the level of absence chosen individually ($v_{12} > 0$), greater strength of relative absence concerns, as captured by an increase in the parameter γ, raises the number of working hours and the duration of absence in market equilibrium, while effort declines.
(c) A greater strength of relative absence concerns has ambiguous consequences for the Pareto-efficient allocation.
(d) The tax rate and level of sick pay which induce a Pareto-efficient allocation as the market outcome are given by

$$0 < t^*\left(c^*, \gamma\right) = -\frac{u_2(c^*, c^*, \gamma)}{u_1(c^*, c^*, \gamma)} < 1 \text{ and } s^*\left(c^*, a^*, \gamma\right) = \frac{\gamma v_2(a^*, \gamma)}{u_1(c^*, c^*, \gamma)}$$

Proof: See Appendix 7.

If there is no absence externality, working hours will be excessive and absence insufficient (cf. Proposition 2). If higher absence by the reference group reduces utility from absence ($v_2 < 0$), there are additional incentives to raise absence. Consequently, it needs no longer to be too low. Therefore, absence externalities indeed mitigate or even compensate for the effect of relative consumption concerns on absence, as surmised in the Introduction. Moreover, the prediction concerning working hours is unaffected by the incorporation of absence externalities. Part (a) of Proposition 3 also clarifies that relative consumption and absence concerns will never balance

out, in that their co-existence induces a Pareto-efficient market equilibrium. This is the case because absence affects utility differently than leisure. In consequence, the composition of effort will never be efficient. Therefore, if relative consumption concerns are complemented by relative absence effects, the efficiency consequences will be fundamentally different than in a setting in which relative consumption and relative leisure effects co-exist. In the latter case, a higher relative consumption level is tantamount to more work and, hence, an inferior relative leisure situation. Such a direct linkage does not exist if preferences exhibit relative absence concerns.

Part (b) of Proposition 3 additionally indicates that the existence of absence externalities (for $v_{12} > 0$) induces individuals to work more and to be absent for a longer duration than in a market outcome without such relative absence effects. However, effort declines. Parts (b) and (c) of Proposition 3 further reveal that there is no straightforward relationship between the strength of relative absence concerns, as measured by the parameter γ, and the efficiency properties of the market outcome. This is the case because both the market equilibrium and the Pareto-efficient allocation change with the strength of relative absence concerns. A rise in the parameter γ increases the marginal utility from absence if $v_{12} > 0$, so that there is more absenteeism. A greater duration of absence raises labour demand and, hence, working hours in equilibrium. As hours are excessive in market equilibrium, a further increase would enlarge the difference between the efficient amount of contractual working time and the market outcome only if the Pareto-efficient allocation were invariant to the strength of the absence externality. With regard to the duration of absence and effort, no such statements are feasible because the absence and effort levels in the market equilibrium may exceed or fall short of their efficient levels.

Finally, Part (d) of Proposition 3 states that the tax rate which induces an efficient choice of working hours is determined by the same ratio of marginal utilities as it is the case in a setting either without absence externality or without the possibility to determine the absence level individually.[12] This structural equality comes about because the tax rate only corrects the distortion in working hours resulting from relative consumption concerns. Part (d) additionally shows that the distortion due to relative absence considerations requires the level of sick pay to be negative if utility decreases with the reference level of absence ($v_2 < 0$), and to be positive otherwise.[13] This is the case because such preferences will induce individuals to choose an excessive level of absence, for a given amount of working hours. Clearly, if sick pay also insured individuals against income variations, an argument in favour of a positive level would arise. Nevertheless, the above analysis indicates that the optimal level of sick pay is reduced by the presence of relative absence concerns.

---

[12] Since the Pareto-efficient consumption level may be higher or lower if there are absence externalities than in a setting without such externalities, the magnitude of $t^*(c^*, \gamma)$, relative to $\hat{t}$, cannot be determined. An exception arises if consumption levels are the same, as it will be true for an iso-elastic utility function u. Since the Pareto-efficient consumption level does not vary with γ in such a setting, tax rates are also the same, i.e. $t^*(c^*, \gamma) = \hat{t}$.

[13] If sick pay were taxed, the level inducing efficient behaviour would have to be higher in absolute terms in order to counteract the mitigating impact of taxes and given by $s^*(c^*, a^*, \gamma) = \gamma v_2(a^*, \gamma)/(u_1(c^*, c^*, \gamma) + u_2(c^*, c^*, \gamma))$.

Going beyond Proposition 3, a setting may be considered in which, for example, the level of sick pay cannot be chosen optimally. Suppose for illustrative purposes that sick pay is too high, $s > s^*(c^*, a^*, \gamma)$. Accordingly, working hours and absence exceed their optimal levels, h^* and a^* (see also Appendix 6). If higher taxes (weakly) raise working hours, the second best optimal tax rate will surely be less than $t^*(c^*, \gamma)$. This will be the case because a reduction in t will mitigate the increases in working hours and in absence that are due to sick pay exceeding its optimal level.

5 Conclusions

In this chapter, we complement a model featuring positional consumption preferences with relative absence concerns. The former externality induces individuals to work too much, the latter has the opposite impact. We show that the net impact on actual working hours, i.e. effort, may coincidentally be zero. However, irrespective of the overall impact on effort, its composition will never be efficient and working hours will always be too high. The main modelling assumption determining this prediction, for which there is also substantial evidence, is that leisure and absence have different utility effects. As a consequence, relative absence concerns do not invalidate but only modify the case for taxation due to positional income considerations. In particular, the tax rate inducing the efficient number of working hours is positive but generally depends on the strength of relative absence concerns. Moreover, relative absence concerns imply that the level of sick pay-inducing efficiency must be less than the amount which is optimal in the absence of such externality. Since the empirical evidence suggests that relative absence concerns focus on colleagues, the optimal level of sick pay may consequently be firm- or even workplace-specific.

The present analysis constitutes a first attempt to model the co-existence of relative absence concerns and positional income preferences. To do so, the investigation relies on several simplifying assumptions, discussed in more detail at the beginning of Sect. 3. Therefore, it may be worthwhile to investigate if the findings derived above will also hold if individuals differ, for example, (a) in the strength of positional preferences, (b) with respect to the other individuals they compare to or (c) the scope for alterations in working hours and absence. Furthermore, we have assumed competitive markets, the absence of unemployment and of adjustments at the extensive margin both for individuals and firms.

Appendix

1. Utility Maximum

From an individual's perspective, the reference levels of consumption and absence, profits, the tax rate and the lump-sum transfer are constant. Therefore, the second-order conditions for a maximum are given by $Z_{hh}, Z_{aa} < 0 < Z_{hh}Z_{aa} - (Z_{ah})^2$, where Z_h and Z_a are defined in Eqs. (3a) and (3b) and the net wage equals $w^N = w[1 - t]$.

$$Z_{hh} = u_{11}\left[w^N\right]^2 - H'' < 0 \tag{15a}$$

$$Z_{aa} = u_{11}\left[s - w^N\right]^2 + v_{11} < 0 \tag{15b}$$

$$Z_{ha} = Z_{ah} = u_{11}w^N\left[s - w^N\right] \geq 0 \tag{15c}$$

Hence, we have

$$Z_{hh}Z_{aa} - Z_{ah}{}^2 = u_{11}\left[v_{11}\left[w^N\right]^2 - H''\left[s - w^N\right]^2\right] - H''v_{11} > 0 \tag{16}$$

2. Stability of Market Equilibrium

In equilibrium, lump-sum payments, T, are determined endogenously in order to balance the budget. Thus, $T = wt[h - a]$. Moreover, profits as defined in (4) are paid out to individuals and affect their consumption. Hence, the equilibrium level of consumption equals production:

$$\begin{aligned} c^m &= w[1-t]\left[h^m - a^m\right] + sa^m + wt\left[h^m - a^m\right] + f(h^m - a^m) - w\left[h^m - a^m\right] - sa^m \\ &= f(h^m - a^m) \end{aligned} \tag{17}$$

To ascertain whether the market equilibrium is stable, we calculate the Jacobian determinant $|J|$ of the system defined by Eqs. (3a), (3b) and (5), taking into account (17). Moreover, all individuals behave identically. Hence, changes in consumption, c, and the reference level, $\overline{c}$, are the same. Similarly, the variations in a and $\overline{a}$ coincide. Thus, the derivatives of (3a), (3b) and (5) with respect to contractual hours, h, absence, a, and wages, w, incorporating (17), are given by $\pi_{hh} = f'' = -\pi_{ha} < 0$, $\pi_{hw} = -1$ and:

$$Z^e_{hh} = [u_{11} + u_{12}]f'w^N - H'' < 0 \tag{18a}$$

$$Z^{e}_{aa} = -[u_{11} + u_{12}]f'\left[s - w^{N}\right] + v_{11} + \gamma v_{12} < 0 \tag{18b}$$

$$Z^{e}_{ha} = -[u_{11} + u_{12}]w^{N}f' > 0 \tag{18c}$$

$$Z^{e}_{ah} = [u_{11} + u_{12}]\left[s - w^{N}\right]f' > 0 \tag{18d}$$

$$Z^{e}_{hw} = u_1[1 - t] = -Z^{e}_{aw} > 0 \tag{18e}$$

In (18a) to (18e), we use the superscript e to indicate that equilibrium repercussions via lump-sum payments T and profits are incorporated. The Jacobian determinant $|J|$ of the system defined by the modified Eqs. (3a), (3b) and (5) is negative.

$$\begin{aligned} |J| &= -\left[Z^{e}_{hh}Z^{e}_{aa} - Z^{e}_{ha}Z^{e}_{ah}\right] - f''Z^{e}_{hw}\left[Z^{e}_{ha} + Z^{e}_{aa} + Z^{e}_{hh} + Z^{e}_{ah}\right] \\ &= -H''\left[[u_{11} + u_{12}]\left[s - w^{N}\right]f' - \left[v_{11} + \gamma v_{12}\right]\right] \\ &\quad + f''u_1[1 - t]\left[H'' - \left[v_{11} + \gamma v_{12}\right]\right] < 0 \end{aligned} \tag{19}$$

3. Pareto-Efficient Allocation

The second-order conditions for a maximum of Γ are

$$\Gamma_{hh} = [u_{11} + 2u_{12} + u_{22}]\left[f'\right]^2 + [u_1 + u_2]f'' - H'' < 0 \tag{20a}$$

$$\Gamma_{aa} = [u_{11} + 2u_{12} + u_{22}]\left[f'\right]^2 + [u_1 + u_2]f'' + v_{11} + 2\gamma v_{12} + \gamma^2 v_{22} < 0 \tag{20b}$$

and $\text{Det} = \Gamma_{hh}\Gamma_{aa} - \Gamma_{ha}{}^2 > 0$. Using

$$\Gamma_{ha} = -\underbrace{[u_{11} + 2u_{12} + u_{22}]}_{(-)}\left[f'\right]^2 - \underbrace{[u_1 + u_2]f''}_{(-)} = -\Gamma_{hh} - H'' > 0, \tag{21}$$

the determinant of the system of Eqs. (8a) and (8b) is found to be positive.

$$\begin{aligned} \text{Det} = &\underbrace{\left[[u_{11} + 2u_{12} + u_{22}]\left[f'\right]^2 + [u_1 + u_2]f''\right]}_{(-)}\underbrace{\left[v_{11} + 2\gamma v_{12} + \gamma^2 v_{22} - H''\right]}_{(-)} \\ &- H''\underbrace{\left[v_{11} + 2\gamma v_{12} + \gamma^2 v_{22}\right]}_{(-)} > 0 \end{aligned} \tag{22}$$

4. Alternative Specification of Preferences

Suppose preferences are given by

$$\tilde{Z}(h,a) = u(c,\overline{c}) - \tilde{H}(h,a) + v(a,\gamma\overline{a}), \tag{23}$$

where $\tilde{H}_1 < 0 < \tilde{H}_2$. Pareto-efficiency can then be characterised by maximising:

$$\tilde{\Gamma}(h,a) = u\big(c(h,a),\overline{c}\big(\overline{h},\overline{a}\big)\big) - \tilde{H}(h,a) + v(a,\gamma\overline{a}) \tag{24}$$

The first-order conditions for individually optimal choices and describing Pareto-efficiency are

$$\tilde{Z}_h = u_1 w[1-t] - \tilde{H}_1(h,a) = 0 \tag{25a}$$

$$\tilde{Z}_a = u_1\big[s - w^N\big] - \tilde{H}_2(h,a) + v_1 = 0 \tag{25b}$$

$$\tilde{\Gamma}_h = [u_1 + u_2]f'(h-a) - \tilde{H}_1(h,a) = 0 \tag{26a}$$

$$\tilde{\Gamma}_a = -[u_1 + u_2]f'(h-a) - \tilde{H}_2(h,a) + v_1 + \gamma v_2 = 0 \tag{26b}$$

The properties of the model will be unaffected if (1) $\tilde{H}_2$ is not too large in absolute value such that (25b) and (26b) define interior solutions for absence choices, and (2) $v_1(a,\gamma\overline{a}) - \tilde{H}_2(h,a)$ exhibits the same qualitative features as $v_1(a,\gamma\overline{a})$ with respect to a.

5. Proof of Proposition 2

Part (a): *If income is untaxed, working hours will be excessive, there will be too little absence from work and, hence, effort and consumption will also be excessive.*

The combination of (11a), (11b), and (12a), (12b) shows that

$$H'(h^m) = v_1(a^m) \tag{27a}$$

and

$$H'\big(h^*\big) = v_1\big(a^*\big). \tag{27b}$$

Given the assumptions on the derivatives (H', H'', $v_1 > 0 > v_{11}$), there are three possible combinations of market outcomes relative to the efficient combination:

Case (1): $h^m = h^*$ and $a^m = a^*$,

Case (2): $h^m < h^*$ such that $H'(h^m) < H'(h^*)$ and $v_1(a^m) < v_1(a^*)$, which implies $a^m > a^*$,

Case (3): $h^m > h^*$ and $a^m < a^*$, according to the same line of argument as in Case (2).

In Case (1), Eqs. (27a) and (27b) hold, but (11a) and (12a), respectively (11b) and (12b), cannot be satisfied simultaneously. Therefore, $h^m = h^*$ and $a^m = a^*$ do not guarantee that the conditions which characterise the market equilibrium and the efficient outcome are both fulfilled.

In Case (2), $h^m - a^m < h^* - a^*$ results, which implies that $c^m = f(h^m - a^m) < c^* = f(h^* - a^*)$ holds. This, in turn, indicates that $f'(h^m - a^m) > f'(h^* - a^*)$ and $u_1(c^m) > u_1(c^*)$ due to the strict concavity of f and u. Furthermore, deducting (12a) from (11a) yields

$$\begin{aligned} u_1(c^m, \bar{c})\, f'(h^m - a^m) - H'(h^m) - \left[\left[u_1(c^*, c^*) + u_2(c^*, c^*)\right] f'(h^* - a^*) - H'(h^*)\right] \\ = \underbrace{u_1(c^m, \bar{c})\, f'(h^m - a^m) - u_1(c^*, c^*)\, f'(h^* - a^*)}_{=\text{A1}} \\ + \underbrace{H'(h^*) - H'(h^m)}_{=\text{A2}} \underbrace{- u_2(c^*, c^*)\, f'(h^* - a^*)}_{(-)} = 0 \end{aligned} \tag{28}$$

In Case (2), the terms A1 and A2 are positive. Therefore, equality (28) cannot hold and $h^m < h^*$, $a^m > a^*$ do not describe the market outcome relative to the efficient situation.

In consequence, the only constellation of working hours and absence which simultaneously guarantees the conditions which describe the market equilibrium and the Pareto-efficient allocation is described by Case (3). If $h^m > h^*$ and $a^m < a^*$, $c^m > c^*$ must also hold. ■

Part (b): *The tax rate* $\hat{t} := t(s = \gamma = 0)$ *which ensures that individuals choose the optimal number of working hours and the optimal duration of absence is given by*

$$0 < \hat{t} = -\frac{u_2(c^*, c^*)}{u_1(c^*, c^*)} < 1. \tag{29}$$

This part can be demonstrated by substituting $\hat{t} := t(s = \gamma = 0) = -u_2(c^*, c^*)/u_1(c^*, c^*)$ in Eqs. (11a) and (11b). Given a unique market equilibrium, it can only be characterised by the values of h and a which fulfil Eqs. (12a) and (12b), i.e. the Pareto-efficient combination. As tax receipts are returned to individuals and they obtain all profit income, consumption will be the same as in the Pareto-efficient allocation, given the same levels of working hours and absence. ■

6. Sick Pay

In market equilibrium, consumption equals production, $c^{\mathrm{m}} = f(h^{\mathrm{m}} - a^{\mathrm{m}})$. Therefore, the derivatives of the first-order conditions (3a), (3b) and (5) with regard to sick pay, s, are $Z^{\mathrm{e}}_{\mathrm{hs}} = \pi_{\mathrm{hs}} = 0$ and $Z^{\mathrm{e}}_{\mathrm{as}} = u_1$. Also taking into account (18a) to (18e), the changes in contractual hours, absence and effort due to a rise in sick pay, s, are found to be

$$\frac{\mathrm{d}h^{\mathrm{m}}}{\mathrm{d}s} = \frac{u_1}{|J|}\left[f'' Z^{\mathrm{e}}_{\mathrm{hw}} - Z^{\mathrm{e}}_{\mathrm{ah}}\right] > 0 \tag{30a}$$

$$\frac{\mathrm{d}a^{\mathrm{m}}}{\mathrm{d}s} = \frac{u_1}{|J|}\left[Z^{\mathrm{e}}_{\mathrm{hh}} + f'' Z^{\mathrm{e}}_{\mathrm{hw}}\right] > 0 \tag{30b}$$

$$\frac{\mathrm{d}(h^{\mathrm{m}} - a^{\mathrm{m}})}{\mathrm{d}s} = -\frac{u_1[u_{11} + u_{12}]f's}{|J|} < 0 \tag{30c}$$

7. Proof of Proposition 3

Notation: Market outcomes in a world with absence externalities are denoted by $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma)$, $a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma)$ and $c^{\mathrm{m}}(h^{\mathrm{m}}, a^{\mathrm{m}}) = c^{\mathrm{m}}(h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma), a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma))$, while the Pareto-efficient allocation is characterised by $h^*(a^*, \gamma)$, $a^*(h^*, \gamma)$, and $c^*(\gamma)$.

Part (a): *If the tax rate is zero ($t = 0$), sick pay is non-negative ($s \geq 0$), and higher absence by the reference group does not raise utility from absence ($v_2 \leq 0$), working hours in market equilibrium will be excessive, while the differences between Pareto-efficient and market outcomes with respect to absence and consumption are indeterminate.*

The comparison of the first-order conditions characterising the market equilibrium and the Pareto-efficient outcome or of a combination of them does not provide insights with respect to the relative levels of working hours and absence. However, it can be shown that only a number of combinations of h, a and $h - a$ are feasible. Basically, the differences $[h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) - a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma)] - [h^*(a^*, \gamma) - a^*(h^*, \gamma)]$, $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) - h^*(a^*, \gamma)$ and $a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma) - a^*(h^*, \gamma)$ could be positive, zero or negative. Hence, the theoretically maximal number of outcomes is 27. To simplify the subsequent argument, note that imposing a sign on the term Diff 1: $= h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) - a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma) - [h^*(a^*, \gamma) - a^*(h^*, \gamma)] = h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) - h^*(a^*, \gamma) - [a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma) - a^*(h^*, \gamma)]$ implies that the same sign applies to the difference Diff 2: $= c^{\mathrm{m}}(h^{\mathrm{m}}, a^{\mathrm{m}}) - c^*(\gamma)$ because $c = f(h - a)$.

Some of the 27 feasible combinations are logically impossible. If Diff 2 > (<) 0 holds, $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) - h^*(a^*, \gamma) \leq (\geq) 0$ and $a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma) - a^*(h^*, \gamma) \geq (\leq) 0$ cannot occur

Table 1 Feasible and impossible combinations of working hours, absence and effort

Sign of Diff 1 & Diff 2	Sign of $h^m(a^m, \gamma) - h^*(a^*, \gamma)$	Sign of $a^m(h^m, \gamma) - a^*(h^*, \gamma)$	Not feasible because
+	+	+	
+	+	0	
+	+	−	
+	0	+	logically impossible
+	0	0	logically impossible
+	0	−	of argument B
+	−	+	logically impossible
+	−	0	logically impossible
+	−	−	of argument B
0	+	+	
0	+	0	logically impossible
0	+	−	logically impossible
0	0	+	logically impossible
0	0	0	of argument A
0	0	−	logically impossible
0	−	+	logically impossible
0	−	0	logically impossible
0	−	−	of argument A
−	+	+	
−	+	0	logically impossible
−	+	−	logically impossible
−	0	+	of argument A
−	0	0	logically impossible
−	0	−	logically impossible
−	−	+	of argument A
−	−	0	of argument A
−	−	−	of argument A

simultaneously. This argument rules out 4 (and another 4) of the 27 combinations. Additionally, if Diff 2 = 0 holds, $h^m(a^m, \gamma) - h^*(a^*, \gamma)$ and $a^m(h^m, \gamma) - a^*(h^*, \gamma)$ must have the same signs. Hence, another six combinations cannot describe the market outcome relative to the Pareto-efficient allocation (cf. Table 1).

We next consider the case of Diff 2 ≤ 0 again. This implies that $u_1(c^m) \geq u_1(c^*)$, given $u_{11} + u_{12} < 0$ and $f'(h^m - a^m) = w \geq f'(h^* - a^*)$. As a result, $u_1(c^m)w > [u_1(c^*) + u_2(c^*)]f'(h^* - a^*)$, since $u_2 < 0$. The comparison of (3a) and (8a), assuming $t = 0$, clarifies that $H'(h^m) > H'(h^*) < 0$ must hold, because otherwise the equations cannot be fulfilled simultaneously. Given the convexity of H in h, $H'(h^m) > H'(h^*)$ implies

that $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) > h^*(a^*, \gamma)$ holds. Accordingly, all theoretically feasible cases for which Diff $2 \leq 0$ is assumed are only compatible with $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) > h^*(a^*, \gamma)$, ruling out a further 6 of the remaining 13 $(27 - 4 - 4 - 6)$ combinations as incompatible with $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma)$, $a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma)$ characterising the market equilibrium and $h^*(a^*, \gamma)$, $a^*(h^*, \gamma)$ the Pareto-efficient allocation (argument A).

Note that thus far the proof has required no restrictions with respect to sick pay and the sign of v_2. Suppose, next, that $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) \leq h^*(a^*, \gamma)$ holds. In accordance with the above line of argument, this implies that $H'(h^{\mathrm{m}}) < H'(h^*)$ is true. Combining (3a), (3b) and (8a), (8b) yields

$$u_1(c^{\mathrm{m}})s - H'(h^{\mathrm{m}}) + v_1(a^{\mathrm{m}}) = 0 \tag{31a}$$

$$\gamma v_2 - H'\left(h^*\right) + v_1\left(a^*\right) = 0 \tag{31b}$$

For $\gamma v_2 < 0$ and $s \geq 0$ or $\gamma v_2 = 0$ and $s > 0$, Eqs. (31a) and (31b) can only hold at the same time if $v_1(a^{\mathrm{m}}) < v_1(a^*)$, that is for $a^{\mathrm{m}}(h^{\mathrm{m}}, \gamma) > a^*(h^*, \gamma)$, and given $v_{11} < 0$ (argument B). Hence, two further combinations have been ruled out. Because no further incompatibilities of the first-order conditions, or combinations thereof, can be discerned, the above considerations leave 5 of the 27 permutations (see Table 1). All of them are characterised by $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) > h^*(a^*, \gamma)$. ■

The proof that $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) > h^*(a^*, \gamma)$ is the only feasible outcome, assumes either a positive level of sick pay ($s > 0$) and $\gamma v_2 \geq 0$, or non-negative sick pay ($s \geq 0$) and envy with respect to absence ($\gamma v_2 < 0$); cf. argument B. Therefore, it also covers the case of positive sick pay and no absence externality. Hence, the above argument constitutes an alternative to the proof provided in Appendix 6 establishing that working hours in market equilibrium will be excessive if sick pay is positive.

Part (b): *If a higher absence level by the reference group increases the level of absence chosen individually ($v_{12} > 0$), a greater strength of relative absence concerns, as captured by an increase in the parameter γ, raises the number of working hours and the duration of absence in market equilibrium, while effort declines.*

Since $Z^e_{\mathrm{h}\gamma} = \pi_{\mathrm{h}\gamma} = 0$ and $Z^e_{\mathrm{a}\gamma} = v_{12}\overline{a}$, the changes in working hours, absence and effort are

$$\frac{\mathrm{d}h^{\mathrm{m}}}{\mathrm{d}\gamma} = -Z^{\mathrm{e}}_{\mathrm{a}\gamma} \underbrace{\frac{Z^{\mathrm{e}}_{\mathrm{ah}} + f'' Z^{\mathrm{e}}_{\mathrm{aw}}}{|J|}}_{(-)} \tag{32a}$$

$$\frac{\mathrm{d}a^{\mathrm{m}}}{\mathrm{d}\gamma} = Z^{\mathrm{e}}_{\mathrm{a}\gamma} \underbrace{\frac{Z^{\mathrm{e}}_{\mathrm{hh}} + Z^{\mathrm{e}}_{\mathrm{hw}} f''}{|J|}}_{(+)} \tag{32b}$$

$$\frac{\mathrm{d}(h^{\mathrm{m}} - a^{\mathrm{m}})}{\mathrm{d}\gamma} = -Z^{\mathrm{e}}_{\mathrm{a}\gamma}\frac{Z^{\mathrm{e}}_{\mathrm{hh}} + Z^{\mathrm{e}}_{\mathrm{ha}}}{|J|} = Z^{\mathrm{e}}_{\mathrm{a}\gamma}\underbrace{\frac{H''}{|J|}}_{(-)} \tag{32c}$$

Part (c): *A greater strength of relative absence concerns has ambiguous consequences for the Pareto-efficient allocation.*

The partial derivatives of Eqs. (8a) and (8b) with respect to γ are given by $\Gamma_{\mathrm{h}\gamma} = 0$ and $\Gamma_{\mathrm{a}\gamma} = v_2 + (v_{12} + \gamma v_{22})\overline{a}$. Since $\Gamma_{\mathrm{a}\gamma}$ cannot be signed without specifying the utility function v, the changes in working hours, absence and effort in the Pareto-efficient allocation are ambiguous. ■

Part (d): *The tax rate and level of sick pay which induce a Pareto-efficient allocation as market outcomes are given by* $0 < t^*(c^*, \gamma) = -\frac{u_2(c^*, c^*, \gamma)}{u_1(c^*, c^*, \gamma)} < 1$ and $s^*(c^*, a^*, \gamma) = \frac{\gamma v_2(a^*, \gamma)}{u_1(c^*, c^*, \gamma)}$.

Replacing t and s in Eqs. (3a) and (3b) by $-u_2(c^*, c^*, \gamma)/u_1(c^*, c^*, \gamma)$ and $\gamma v_2(a^*)/u_1(c^*, c^*, \gamma)$ and using $w = f'(h - a)$ from (5) shows that Eqs. (3a) and (3b) will hold for those values of working hours and absence which characterise the Pareto-efficient allocation described by Eqs. (8a) and (8b). All tax payments are returned to individuals via lump-sum payments. Moreover, individuals obtain the entire profit income. Consequently, income and consumption will be the same as in the Pareto-efficient allocation, given $h^{\mathrm{m}}(a^{\mathrm{m}}, \gamma) = h^*(a^*, \gamma)$ and $a^{\mathrm{m}} = a^*$. ■

References

Alpizar, F., Carlsson, F., & Johansson-Stenman, O. (2005). How much do we care about absolute versus relative income and consumption? *Journal of Economic Behavior & Organization, 56*(5), 405–421.

Alvarez-Cuadrado, F. (2007). Envy, leisure, and restrictions on working hours. *Canadian Journal of Economics, 40*(6), 1286–1310.

Aronsson, T., Blomquist, S., & Sacklén, H. (1999). Identifying interdependent behaviour in an empirical model of labour supply. *Journal of Applied Econometrics, 14*(4), 607–626.

Aronsson, T., & Johansson-Stenman, O. (2010). Positional concerns in an OLG-model: Optimal labor and capital income taxation. *International Economic Review, 51*(4), 1071–1095.

Aronsson, T., & Johansson-Stenman, O. (2013). Conspicuous leisure: Optimal income taxation when both relative consumption and relative leisure matter. *The Scandinavian Journal of Economics, 115*(1), 155–175.

Aronsson, T., & Johansson-Stenman, O. (2018). Paternalism against Veblen: Optimal taxation and non-respected preferences for social comparisons. *American Economic Journal: Economic Policy, 10*(1), 39–76.

Arrow, K. J., & Dasgupta, P. S. (2009). Conspicuous consumption, inconspicuous leisure. *The Economic Journal, 119*(541), F497-516.

Boskin, M. T., & Sheshinski, E. (1978). Optimal redistributive taxation when individual welfare depends upon relative income. *The Quarterly Journal of Economics, 92*(4), 589–601.

Bracha, A., Gneezy, U., & Loewenstein, G. (2015). Relative pay and labor supply. *Journal of Labor Economics, 33*(2), 297–315.

Bradley, S., Green, C., & Leeves, G. (2007). Worker absence and shirking: Evidence from matched teacher-school data. *Labour Economics, 14*(5), 319–334.
Bradley, S., Green, C., & Leeves, G. (2014). Employment protection, threat and incentives effects on worker absence. *British Journal of Industrial Relations, 52*(2), 333–358.
Braham, M., & Steffen, F. (Eds.). (2008). *Power, freedom, and voting—Essays in honour of Manfred J. Holler*. Springer-Verlag.
Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics, 133*(2), 611–663.
Cahuc, P., & Postel-Vinay, F. (2005). Social status and the overworked consumer. *Annales d'Économie et de Statistique, 78*, 143–161.
Carrieri, V. (2012). Social comparison and subjective well-being: Does the health of others matter? *Bulletin of Economic Research, 64*(1), 31–55.
Carlsson, F., Johansson-Stenman, O., & Martinsson, P. (2007). Do you enjoy having more than others? Survey evidence of positional goods. *Economica, 74*(No. 296), 586–559.
Choudhary, M. A., & Levine, P. (2006). Idle worship. *Economics Letters, 90*(1), 77–83.
Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature, 46*(1), 95–144.
Corneo, G. (2002). The efficient side of progressive income taxation. *European Economic Review, 46*(7), 1359–1368.
Dale-Olsen, H., Østbakken, K. M., & Schøne, P. (2015). Imitation, contagion, or exertion? Using a tax reform to reveal how colleagues' sick leave influences worker behaviour. *The Scandinavian Journal of Economics, 117*(1), 57–83.
De Paola, M. (2010). Absenteeism and peer interaction effects: Evidence from an Italian Public Institute. *The Journal of Socio-Economics, 39*(5), 420–428.
Dodds, S. (2012). Redistributive taxation with heterogeneous relative consumption concerns. *Canadian Journal of Economics, 45*(1), 220–246.
Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology, 29*(1), 94–122.
Duesenberry, J. S. (1949). *Income, savings and the theory of consumer behavior*. Harvard University Press.
Dupor, B., & Liu, W.-F. (2003). Jealousy and equilibrium overconsumption. *American Economic Review, 93*(1), 423–428.
Eckerstorfer, P. (2014). Relative consumption concerns and the optimal tax mix. *Journal of Public Economic Theory, 16*(6), 936–958.
Fisher, W. H., & Hof, F. X. (2008). The quest for status and endogenous labor supply: The relative wealth framework. *Journal of Economics, 93*(2), 109–144.
Frank, R. H. (1985). The demand for unobservable and other nonpositional goods. *American Economic Review, 75*(1), 101–116.
Frick, B., & Malo, M. Á. (2008). Labor market institutions and individual absenteeism in the European Union: The relative importance of sickness benefit systems and employment protection legislation. *Industrial Relations, 47*(6), 505–529.
Godøy, A., & Dale-Olsen, H. (2018). Spillovers from gatekeeping—Peer effects in absenteeism. *Journal of Public Economics, 167*, 190–204.
Goerke, L. (2019). Absence from work, sick pay and positional consumption concerns. *Labour: Review of Labour Economics and Industrial Relations, 33*(2), 187–211.
Goerke, L., & Hillesheim, I. (2013). Relative consumption, working time and trade unions. *Labour Economics, 24*, 170–179.
Goerke, L., & Holler, M. J. (1997). *Arbeitsmarktmodelle*. Springer-Verlag.
Goerke, L., & Neugart, M. (2021). Social preferences, monopsony, and government intervention. *Canadian Journal of Economics, 54*(2), 864–891.

Gómez, M. A. (2008). Consumption and leisure externalities, economic growth and equilibrium efficiency. *Scottish Journal of Political Economy, 55*(2), 227–249.
Grolleau, G., & Saïd, S. (2008). Do you prefer having more or more than others? Survey evidence on positional concerns in France. *Journal of Economic Issues, 42*(6), 1145–1158.
Hansen, J. (2000). The effect of work absence on wages and wage gaps in Sweden. *Journal of Population Economics, 13*(1), 45–55.
Heffetz, O. (2011). A test of conspicuous consumption: Visibility and income elasticities. *The Review of Economics and Statistics, 93*(6), 1101–1117.
Hesselius, P. (2007). Does sickness absence increase the risk of unemployment? *The Journal of Socio-Economics, 36*(2), 288–310.
Hesselius, P., Johansson, P., & Nilsson, J. P. (2009). Sick of your colleagues' absence? *Journal of the European Economic Association, 7*(2–3), 583–594.
Hesselius, P., Johansson, P., & Vikström, J. (2013). Social behaviour in work absence. *The Scandinavian Journal of Economics, 115*(6), 995–1019.
Holler, M. J. (1986). *Ökonomische Theorie des Arbeitsmarktes*. Wissenschaftliche Buchgesellschaft.
Ichino, A., & Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large Italian firm. *The Quarterly Journal of Economics, 115*(5), 1057–1090.
Ireland, N. J. (1998). Status-seeking, income taxation and efficiency. *Journal of Public Economics, 70*(1), 99–113.
Johansson, P., Karimi, A., & Nilsson, J. P. (2019). Worker absenteeism: Peer influences, monitoring, and job flexibility. *Journal of the Royal Statistical Society A, 182*(Part 2), 605–621.
Kuhn, P., Kooreman, P., Soetevent, A., & Kypteyn, A. (2011). The effects of lottery prizes on winners and their neighbors: Evidence from the Dutch postcode lottery. *American Economic Review, 101*(5), 2226–2247.
Lindbeck, A., Palme, M., & Persson, M. (2016). Sickness absence and local benefit cultures. *The Scandinavian Journal of Economics, 118*(1), 49–78.
Liu, W.-F., & Turnovsky, S. J. (2005). Consumption externalities, production externalities and long-run macroeconomic efficiency. *Journal of Public Economics, 89*(5–6), 1097–1129.
Ljungqvist, L., & Uhlig, H. (2000). Tax policy and aggregate demand management under catching up with the Joneses. *American Economic Review, 90*(5), 356–366.
Markussen, S. (2012). The individual cost of sick leave. *Journal of Population Economics, 25*(6), 1287–1306.
Miraglia, M., & Johns, G. (2021). The social and relational dynamics of absenteeism from work: A multilevel review and integration. *Academy of Management Annals, 15*(1), 37–67.
MISSOC (Mutual Information System on Social Protection). (2021). *Comparative tables database—Sickness cash benefits*. Retrieved December 15, 2021.
Mujcic, R., & Frijters, P. (2015). Conspicuous consumption, conspicuous health, and optimal taxation. *Journal of Economic Behavior & Organization, 111*, 59–70.
Neumark, D., & Postlewaite, A. (1998). Relative income and the rise in married women's employment. *Journal of Public Economics, 70*(1), 157–183.
OECD. (2010). *Sickness, disability, and work: Breaking the barriers—A synthesis of findings across OECD countries*, Paris.
Palme, M., & Persson, M. (2020). Sick pay insurance and sickness absence: Some European cross-country observations and a review of previous research. *Journal of Economic Surveys, 34*(1), 85–108.
Park, Y. (2010). The second paycheck to keep up with the Joneses: Relative income concerns and labor market decisions of married women. *Eastern Economic Journal, 36*(2), 255–276.
Pérez-Asenjo, E. (2011). If happiness is relative, against whom do we compare ourselves? Implications for labour supply. *Journal of Population Economics, 24*(6), 1411–1442.
Persson, M. (1995). Why are taxes so high in egalitarian societies? *The Scandinavian Journal of Economics, 97*(6), 569–580.

Pichler, S., & Ziebarth, N. R. (2017). The pros and cons of sick pay schemes: Testing for contagious presenteeism and noncontagious absenteeism behavior. *Journal of Public Economics, 156*, 14–33.
Pingle, M., & Mitchell, M. (2002). What motivates positional concerns for income? *Journal of Economic Psychology, 23*(1), 127–148.
Schor, J. (1991). *The overworked American: The unexpected decline of leisure.* Basic Books.
Scoppa, V., & Vuri, D. (2014). Absenteeism, unemployment and employment protection legislation: Evidence from Italy. *IZA Journal of Labor Economics, 3*, 3.
Seidman, L. S. (1988). The welfare costs of a relativistic economy. *Journal of Post Keynesian Economics, 11*(2), 295–304.
Skåtun, D. R., & Skåtun, J. D. (2004). The impact of turnout on turning up: The complementarity of attendance among co-workers. *Journal of Economics, 83*(5), 225–242.
Solnick, S. J., & Hemenway, D. (2005). Are positional concerns stronger in some domains than in others? *American Economic Review P & P, 95*(2), 147–151.
Wendner, R. (2014). Ramsey, Pigou, heterogeneous agents, and nonatmospheric consumption externalities. *Journal of Public Economic Theory, 16*(3), 491–521.
Wouters, S., van Exel, N., Job, A., van de Donk, M., Rohde, K. I. M., & Brouwer, W. B. F. (2015). Do people desire to be healthier than other people? A short note on positional concerns for health. *European Journal of Health Economics, 16*(1), 47–54.

Power, Responsibility and Social Policy: The Impact of Basic Income in a Competitive Experimental Labor Market

Veera Jokipalo and Katri Sieberg

1 Introduction

Labor markets are a simple illustration of a traditional market, in which there is supply of a certain type of labor, demand for it, and the intersection between these results in the going (local) price for labor—the wage. As Holler and Leroch (2010) note, however, these types of market transactions can have differing implications in terms of economic efficiency and in terms of justice. "…economists speak of involuntary unemployment and worry about efficiency, while those who experience involuntary unemployment experience asymmetry in how the economy treats them." (2010, 311) A meaningful question, then, is: is there a labor market policy that is efficient and also 'fair'? Such a policy would encourage responsible labor behavior, in terms of effort, and would also minimize costs from problems such as involuntary unemployment.

Basic Income (BI) has, at least on the surface, the potential to meet these criteria. BI can be considered 'fair,' or at least equity increasing, because it redistributes wealth more evenly than what would be seen absent the policy. BI is a policy proposal involving unconditional, universal cash payments, usually pitched at country-level to include all permanent residents. While the potential impact of BI on the labor market is of high interest to politicians, researchers and activists alike, the impact of BI on both equity and on responsibility (i.e. on wages and on effort) is a contested issue even among researchers. For example, some would suggest that the payment could be viewed by employers as a wage subsidy, resulting in lower wage levels. Others stress the potential of BI to act as a strike fund fueling collective action, or a safety net that enables individual workers to demand better wages lest they quit their jobs (Calnitsky, 2018; Widerquist, 2013). This potential could potentially mean that workers would have power that would undermine that of employers. The

V. Jokipalo · K. Sieberg (✉)
Tampere University, Tampere, Finland
e-mail: katri.sieberg@tuni.fi

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_5

impact of BI on effort levels is further debated. It is unclear whether access to a basic level of income will decrease responsibility in terms of decreasing effort, whether it would increase it in jobs that are attractive to employees, or whether effort levels would stay the same. Would the potential for workers to keep BI as they become employed, increase responsibility? A noted problem with unemployment benefits is that they discourage employment because workers immediately lose benefits when they become employed. This problem does not occur with BI. It is important to see whether wages or effort levels would increase or decrease, in order to estimate further effects on labor supply and thus the whole labor market. Maintaining a critical mass of labor market participation is also necessary to keep an income tax funded BI policy afloat, meaning that estimating such effects before implementation is crucial.

Holler and Leroch (2010) note the benefits of experimental economics in highlighting our understanding of the extent to which fairness concerns can affect behavior. We argue that in BI research, economic experiments have the additional benefit of being a much-needed complement to results from field experiments. This is due to the nature of the hypothesized effects of BI, many of which emerge only when BI is implemented both long-term and for all members of the relevant community (Calnitsky, 2019; Widerquist, 2018). By default, a field experiment cannot fulfill both conditions, as this would call for a full-scale implementation of BI. Laboratory experiments come with the caveat of observing people in artificial circumstances but offer instead a way to efficiently control for auxiliary variables, and to run several experiments quickly and cost-effectively to measure the impact of specific changes in the economic environment. The use of laboratory experiments in BI research has previously been argued for in Noguera & De Wispelaere (2006); Jokipalo (2019); and Füllbrunn, Delsen, & Vyrastekova (2019).

This paper reports the results from laboratory experiments run in Irvine, California in February 2020. The experimental design used is based on a study by Jokipalo (2019), who studied the impact of BI on gift exchange behavior using a modified gift exchange game (the original game having been designed by Fehr et al., 1993) with real effort (real effort task coded and freely shared by Benndorf, Rau, & Sölch, 2018), real leisure (see Corgnet et al., 2015), and a repeated game format with partner matching. The novelty in our research is in adding competition into the design in the form of both (1) an oversupply of labor and (2) the possibility for employers and employees to form a labor contract with any employee/employer in the session. Reputation effects are allowed through giving participants fixed screen names.

Our primary research interest is on the impact of BI on wages and on the effort exerted by employees. Here, effort is a measure of *productivity*; it corresponds specifically to how the employee contributes to the product of the firm instead of, for example, the hours worked. Even though the process of wage determination is often simplified to involve working hours only, there is a long line of literature in labor economics describing *efficiency wages*. Efficiency wage theories suggest that employers attempt to minimize labor costs in terms of how much is produced, not by how many man hours their employees perform. The idea was first formally introduced by Alfred Marshall, and since then efforts have been made to understand why and how efficiency wages seem to emerge in some labor markets or in

some interactions between individuals but are not ubiquitous (Marshall, 1925; see also Doering & Piore, 1971; Akerlof, 1984; Yellen, 1984). It is important to know whether the potential increasing impact of BI on wages comes with or without an increase in productivity. This information will help us to understand whether a BI policy would have a negative impact on worker responsibility. In Jokipalo (2019), effort was increased with a BI policy.

2 Investigating the Impact of Basic Income in Experimental Labor Markets

In recent decades, Basic Income has become a widely used umbrella term for a variety of proposed policies. It could be debated that many such proposals are not true BI policies, and that many which are introduced under different names could be identified as BI policies. For the purposes of this paper, we will rely on the definition provided by Basic Income Earth Network (BIEN). According to the organization, BI is a "periodic cash payment unconditionally delivered to all on an individual basis, without means test or work requirement".

In terms of economic and labor market impact, all these features of the policy proposal as well as their precise interpretations in policy design are significant. Periodic payments, as opposed to lump sums, paid indefinitely to recipients mean that people can plan their lives and budgets based on a steady stream of future income. Cash, as opposed to in-kind transfers such as food and housing, enable recipients to freely decide how to make use of the transfer. Individual payments—likely coupled with individual taxation—somewhat separate household labor supply decisions from one another. The lack of a means test allows for transfers to be made to those with savings and other assets, meaning there is no need to liquidate such assets in the case of temporarily decreased cash flow, and the lack of work requirement makes way for considerable flexibility in entering and exiting the labor market.

In this paper, we are especially interested in the labor market effects of indefinite, periodic payments being paid without work requirements. Technically, this would allow recipients to permanently exit the labor force and live on the BI transfer if the transferred amount (possibly combined with personal savings or various forms of non-labor income) were high enough. Even if such a radical move were not feasible, the flexibility to enter and exit the labor market could be used to the advantage of employees, giving them more power than they had without BI. We wish to explore whether this would result in less total effort exerted by employees at their jobs (productivity), lower participation in the labor force (labor supply), or a different division of the value created through labor (wages).

2.1 Results From the Field and From the Lab

Relying on theory alone, there is no one answer to how labor market would change with the implementation of BI. A simple assessment of labor supply and non-labor income would lead us to conclude that if income is provided through another source than labor, and labor is by definition undesirable, labor supply would decrease. This should increase wages to preserve market equilibrium. However, there are other possible outcomes. To mention a few, BI transfers could be seen as a form of wage subsidy, making it more acceptable to pay very low wages. As for productivity, some might choose to work part-time, which may make them more productive during the time they are at work. Moreover, some might use the BI payment to acquire education that helps them become more productive at work, or resign from current positions to find a job where they are more motivated or better suited to perform the work. An increase in productivity increases the benefits employers gain in exchange for wages, having a similar effect to a wage decrease for them.

At the time of writing, there is one generally accepted example of the policy being in place: the Alaska Oil Dividend. Permanent residents of Alaska are eligible to apply for a lump sum in cash each year according to the state's oil revenue. Historically, the payment has ranged from around 300–2000 USD. It should be noted that the yearly sum is lower than the level BI is usually pitched at.[1] Higher level BI has only been experimentally tested, albeit field experiments on the policy have become quite popular in the past few years with various studies being run or planned in countries such as Wales, the US, and Kenya at the time of writing.

After decades of field experiments, empirical evidence on the impact of BI on the labor market is somewhat conflicting. Looking at all the available evidence together, the hypothesis that BI would decrease labor supply is not supported (a meta-analysis was recently reported in Gilbert, Murphy, Stepka, Barrett, & Worku, 2018). However, the results are not identical for all countries (or, equivalently, all experimental designs) or for all recipients. For example, labor supply may increase for those who can newly afford to buy a vehicle required for work, and it may decrease for those currently employed but also heavily involved in unpaid care work. Importantly, not many experiments to date have involved a saturated design where a large community as a whole received BI payments, which could lead to dynamic effects over time as employers react to the policy and any following changes to employees' behavior in general, looking at employees' labor supply decisions without knowing how the chain reaction with employers' wage decisions is only half the equation at best.

It is also important to keep in mind what sort of policy BI is to be compared against. While BI is paid to all, most of its immediate benefits are reaped by those with low

[1] As noted, the size of the lump sum and the purpose for its existence distinguish the Alaska Oil Dividend from a BI policy. In particular, a BI policy has a goal in establishing a base income for a population, whereas the Oil Dividend provided extra income to the population. As an anonymous reviewer noted, the oil dividend could allow a middle class family to use the 2000 USD to buy a motorcycle for their son. In a BI situation, the extra income would be generally taxed away for a middle class family.

income—such as the sick, unemployed, and those performing unpaid labor. It is therefore meaningful to compare BI with existing and proposed schemes designed to alleviate poverty. BI lies at one extreme in the sense that it involves cash payments to all. At the other extreme lies the complete lack of formal income transfers for the poor. Somewhere in between lie the policies in use in most Western countries; in-kind transfers and means-tested cash transfers as well as transfers with working requirements.

As noted, the expected effects of a BI policy are not straightforward. Kolm & Tonin (2015) show that means tested policies can have a beneficial effect on employment and education rates precisely because recipients must be at work to receive them. Their results initially seem to imply that a BI policy would have potentially negative effects. The authors state, however, that they do not explore the effect of unconditional benefits. A policy such as BI, which allows individuals to obtain benefits similar to those gained with unemployment, without discouraging people from working in part-time or low-paying jobs, is somewhat tangential to their model and could have interesting implications.

All in all, we argue that economic experiments can act as a valuable complement to field experiments in gathering data about the effects of a BI policy. Here, we consider two earlier contributions where the impact of a Basic Income on an experimental labor market is explored: Jokipalo (2019) and Haigner, Höchtl, Jenewein, Schneider& Wakolbinger 2012. In both papers, the experimental design involved real-effort tasks used to gauge the labor supply of participants. Both designs also featured a 'Basic Income' treatment which was compared against a control with no simulated cash transfers. There are also several significant differences in the designs used by Jokipalo and Haigner et al. Jokipalo's experiment is a modified gift exchange game with a repeated design and fixed pairings of employers and employees. A 'BI transfer' was paid to each employee in every round in the BI treatment. Employees had the choice between 'real leisure' (see Corgnet et al., 2015) and real-effort during a designated working phase. In contrast, in Haigner et al., only one prolonged 'round' was played in each session. All participants played the role of employees who were paid a fixed piece rate for their real-effort tasks. In the BI treatment, wages were taxed, and the tax proceedings were used to finance a BI transfer for all participants.

Jokipalo and Haigner et al. both conclude that labor supply was not diminished by the BI treatment. In the former paper, it was found that effort levels as well as wages were in fact higher than in the control. In the latter, no significant difference in effort was observed. This may be due to Jokipalo's design allowing for gift exchange behavior, meaning that employees could potentially expect to earn higher wages if they increased their effort levels. The repeated design with fixed pairings may also have particularly encouraged gift exchange behavior.

Notably, neither design featured any form of competition between employees or employers. As competition is both an element present in most real-world labor markets and an important determinant of behavior, we wish to see whether the results of Jokipalo and Haigner et al.—that labor supply would not be decreased by a BI policy—are reproducible in an environment with competition. In addition to employer/employee pairs no longer being fixed, but rather determined through a

wage auction, an oversupply of labor will be introduced. Otherwise, we draw largely from Jokipalo (2019) in our experimental design. A design involving participants assigned as both employers and employees is critical for the assessment of both wages and effort and the way they interact over time. We also appreciate participants being able to flexibly adjust their labor supply through both deciding whether to engage in paid work as well as completing only the desired number of tasks.

2.2 Competition in Gift Exchange Games

An experimental tool frequently used to assess labor market interactions is the gift exchange game. Typical gift exchange game experiments seek to establish the potential for a gift exchange, or the lack thereof. In other words, they examine situations in which employers offer wages to workers when the quality of effort exerted by the worker cannot be determined in advance. If the worker repays the 'gift' of a high wage with a correspondingly high level of effort, the pair have engaged in gift exchange. In order to isolate true gift exchange behavior from behavior resulting from reputation effects—high effort levels given in hopes of maintaining a continued employment relationship—most of the experiments test behavior in situations in which either employers and employees are anonymous and cannot be recognized by each other, or they examine situations without the possibility of repeated play. In general, the bulk of the literature finds that gift exchange behavior can occur. Because our focus is on the effect that BI might have on wages and effort levels in a general employment market, we relax some of the conditions imposed in the literature. We allow for repeated interactions between employers and employees.

As noted, a large portion of the literature finds a potential for gift exchange behavior in laboratory experiments. Fehr, Kirchsteiger, and Riedl (1998a, 1998b) note that previous experiments had failed to identify reciprocity between employers and employees because they had misrepresented the problem. They note that while in some experiments—ultimatum games for example—there is a potential for fairness concerns in distributions of a cake, in more competitive games this tendency decreases substantially. In, for example, experiments testing potential reciprocity in a competitive market, fairness concerns seem to disappear, and buyers and sellers seek the best possible prices for themselves. The authors note that this scenario fails to represent the possibility for reciprocity because the quality of the goods traded is predetermined by the experimenter. Thus, the players can only try to maximize personal payoff. To correct this problem, the authors allow sellers to determine the quality of the good after agreeing on a price. Game theory would predict a low quality for any price offered, thus, low prices should be offered. Any price above the minimum level is, thus, a 'gift,' and any quality level above the minimum is an indication of reciprocity.

Fehr, Kirchsteiger, and Riedl (1998a, 1998b) conduct competitive experiments in which sellers (employees) and buyers (employers) are paired anonymously in an auction. They compare results from treatments in which the sellers had the option

of choosing quality levels (reciprocity treatments—RT) with those in which the quality level was predetermined (control treatment—CT). They find that subjects were highly competitive in the control treatment. They find, in contrast, evidence of gift behavior and reciprocity from the same subjects when they took part in the reciprocity treatment.

Fehr et al. (1998a) have also performed comparisons between competitive and noncompetitive market designs in gift exchange games. They find that if employees are unable to change their effort level, gifts in the form of high wages are not paid. Contrarily, high wages and high effort are observed both with a design in which employers and employees are randomly matched, and where matches are determined through an auction. With the latter design, gift exchange behavior from the employer side is at first diminished compared to the former, but this difference disappears after approximately four experimental rounds. Based on these results, it seems that gift exchange behavior takes place as long as reciprocity is enabled through the design; the possibility of repeated encounters with the same person is not necessarily required. However, List (2006) and Van der Heijden, et al. (2001) have found that the promise of future encounters increases the chance of forms of exchange requiring mutual trust.

Following initial experiments, the gift exchange game has been useful in assessing the effects of various labor conditions or policies. Brandts and Charness (2004) also examine the potential for gift exchange behavior in a market with competition and unemployment. They vary the groups experiencing competition, such that in some treatments they have an oversupply of labor and in others, they have an oversupply of firms. The authors further test the impact of a minimum wage on gift exchange behavior. As in the Fehr, Kirchsteiger, and Riedl (1998a, 1998b) paper, firms made offers through a one-sided auction, and workers had the opportunity to accept. Workers then chose effort levels. Their interests lay in seeing how effort levels changed with a change in the direction of competition—i.e. more competition between workers for jobs or more competition between firms for workers. A higher wage in the first scenario is a greater indication of a gift than in the second, where higher wages can merely reflect competitive pressures. Similarly, they examined the effect of what was perceived 'generous' in the case of minimum wage rules. They found that increase in effort corresponded to increases in wages. They found that wages, in the case of excess supply of labor (ESL) tended to be lower than those in which there was excess supply of firms (ESF) and were distinctly lower than wages accepted in the ESF treatments. They also found that there were fewer wage offers in the high range in the case of minimal wage. Effort levels were also impacted by the minimum wage treatments. The authors speculate that knowing that firms were required to offer at least a certain level of wage made workers less responsive to wages offered (Brandts & Charness, 2004 704).

The goal of our experiments is to test the effects of a BI policy, in contrast to those of no social insurance and unemployment insurance, on wages and effort. We are aware that a laboratory experiment cannot give exact information regarding the real-world effect of a BI policy, as has been noted in the literature (Holler & Leroch, 2010; Holler et al., 1992) experiments are valuable, especially in terms of

policy, in highlighting factors that can shape decisions. Frequently, these factors are ignored in theoretical models. Theoretically, unemployment insurance will lead to an increase in wages (Shapiro & Stiglitz 1984). This insight has been supported by quasi-experiments using French unemployment data (Le Barbanchon, Rathelot, & Roulet 2019). It is, as noted, less clear what effects a basic income policy may have on wages or effort. Brandts and Charness' (2004) results regarding wages and effort in the minimum wage case suggest that a policy in which workers receive a sum of money, outside of the wage, regardless of whether or not they work, may dampen both wage offers and effort. However, because the money is provided by an external agent, rather than the employers being required to award at least a certain wage, the potential impact is less obvious.

In our design, participants also receive information about the general wage level in the market. This allows participants to compare their wage offers to the average wage level in the market. Thus, behavior could in theory converge toward a competitive equilibrium. Based on results from earlier contributions in the literature, it is likely that this design detail will increase both effort and offers, compared to Jokipalo (2019). Siang, Requate, and Waichman find that the addition of wage comparisons increases both effort and wages in repeated relationships, while the result for one-shot games is the opposite (2011). More generally, Fehr and Schmidt note that payoff comparisons in experiments slows down convergence to the competitive equilibrium, perhaps because of inequality aversion (1999). Based on such results, it seems that wage comparisons can help the market reach an equilibrium with high wages and high effort rather than market-clearing wages and minimal effort.

3 Modeling the Effect of Basic Income on Wages and Effort Under Competition

Theoretically, we consider two efficiency wage models, the fair wage-effort hypothesis (Akerlof & Yellen, 1990) and the shirking model (Shapiro & Stiglitz, 1984), as explanations for behavior. These models are not fully consistent with the setup in our experiments, but they provide motivation for what we might expect.[2] In both models, the employee considers payments made to the unemployed when deciding

[2] In particular, the models differ in their approach to unemployment, because the Akerlof version is based on incentivized wage contracts which involve the opportunity to fire workers for not performing at a desired level, while in a gift exchange, normally workers cannot be fired. In our experiment, workers cannot be 'fired' during a round, but they risk unemployment in subsequent rounds, if they do not provide enough effort. The effort, in this case, as in some of the other experiments described above, isn't perfectly a 'gift' because it is reputation building. We selected our design to match labor market situations as closely as possible, but we recognize that some of the predictions from the theoretical models may be less than precise in our case. While the experiments are not a perfect test of either theory, we find it important to include them to provide the theoretical background for the study.

how much effort to exert at work. In the fair wage-effort hypothesis, effort is dependent on whether the worker's wage is lower than what is perceived as the 'fair wage'; fair wage perceptions depend, among other factors, on the unemployment rate and benefits paid to the unemployed. In the shirking model, employers pay a higher wage to encourage effort. This wage leads to decreased demand for labor, creating unemployment. The lack of full employment 'scares' workers into working hard, as they do not want to lose their job, and this can be exacerbated by low benefits for the unemployed. If employers predict this behavior, they will pay higher wages when benefits for the unemployed are high or unemployment is low and vice versa.

As noted, Shapiro & Stiglitz (1984) examine the effect of wages above the market-clearing level and of competition on tendencies to shirk. Their model has an economy of N identical workers who dislike labor but enjoy consumption. The workers' utility over wage, w and effort, e is U(w,e) = w − e. They further assume that effort is either 0 or some fixed level, e > 0. Unemployed workers get a benefit $\overline{w}$ and provide no effort. There is a probability, b, that a worker will lose his job, and a discount factor r. Workers select an effort level, and there is a probability q that shirking behavior will be discovered, and the worker will be fired.

Workers select effort levels to maximize their expected utility. More specifically, Shapiro and Stiglitz compare the discounted utility for an employed shirker with that of an employed non-shirker. They derive the wage and unemployment benefits necessary to make the utility of non-shirking higher than those of shirking.

They state that the equation for a shirker is

$$r V_e^s = w + (b+q)\left(V_u - V_e^s\right)$$

And for a non-shirker it is

$$r V_e^n = w - e + (b)\left(V_u - V_e^n\right)$$

These can be rewritten as

$$V_e^s = \frac{w + (b+q)V_u}{r+b+q}$$

And

$$V_e^n = \frac{w - e + (b)V_u}{r+b}$$

The worker will refrain from shirking only if $V_e^n \geq V_e^S$. . Thus, the No Shirking Condition (NSC) is when

$$\hat{w} \equiv w \geq r V_u + \frac{e}{q}(r+b+q) \tag{1}$$

Implications of their model include the following (Shapiro & Stiglitz, 1984, p. 436):

The critical wage will be higher.

- The higher the required effort level
- The higher the unemployment benefits
- The lower the probability, q, of being caught shirking
- The higher the interest rate
- The higher the probability of losing one's job

Our adaption, using BI, means that a worker receives BI both when employed and when unemployed. The original equation facing all employed workers is:

$$V_e = wt + (1 - rt)[bt(V_u) + (1 - bt)(V_e)]$$

Including BI, (B) we now have:

$$V_e = (w + B)t + (1 - rt)[bt(V_u + B) + (1 - bt)(V_e + B)]$$

This simplifies to

$$V_e = \frac{wt + Bt + btV_u - bt^2rV_u + B - rtB}{rt + bt - rbt^2}$$

Which, with continuous time ($t \to 0$) is

$$V_e = \frac{w + B + bV_u + B - rB}{r + b}$$

The equations for the shirker versus non-shirker are now:

$$rV_e^s = w + B + B(1 - r) + (b + q)\left(V_u - V_e^s\right)$$

And

$$rV_e^n = w + B - e + B(1 - r) + b\left(V_u - V_e^n\right)$$

So

$$V_e^s = \frac{w + B + B(1 - r) + (b + q)V_u}{r + b + q}$$

$$V_e^s = \frac{w + B + B(1 - r) + (b + q)V_u}{r + b + q}.$$

And

$$V_e^n = \frac{w + B - e + B(1-r) + (b)V_u}{r+b}$$

These can be simplified to the following NSC with BI:

$$\breve{w} \equiv w \geq rV_u + B(r-2) + \frac{e}{q}(r+b+q) \tag{2}$$

Rearranged, to take into account that BI is paid in every stage, regardless of income status, we have:

$$\breve{w} + B \equiv w + B \geq rV_u - B(1-r) + \frac{e}{q}(r+b+q)$$

This is nearly identical to Eq. (1) from Shapiro and Stiglitz. It merely includes BI in the wage.

In other words, the wage level, under BI, that is necessary to deter shirking is lower than the wage level with simple unemployment.

Given the theoretical results, we make the following hypotheses[3]:

H1. The highest average wage offers will be observed in the unemployment benefit (UB) treatment.
H2. The UB treatment will have the highest offer size required to induce effort levels higher than the minimum (of one unit of effort in the case of the experiment).
H3. The lowest average wage offers will be observed in the BI setting.
H4. The control setting will have the lowest offer size required to induce effort levels higher than the minimum.

As noted, we also draw from the Fair Wage-effort hypotheses proposed by Akerlof and Yellen (1990). This hypothesis states that workers will compare their wages to what is seen as the fair wage. Their resulting, normalized effort level will be

$$\mathrm{e} = \min(\mathrm{w}/\mathrm{w}*,\ 1)$$

In other words, when the wage is below the fair wage, workers will devote a fraction of the normal effort. They note that "Unemployment occurs when the fair wage w* exceeds the market-clearing wage." (Akerlof & Yellen, 1990, 256). The derivation of the fair wage is somewhat imprecise. The authors propose two possibilities. It can be determined exogenously (Akerlof & Yellen, 1990, 267) or it can be relative to the wages from a reference group. The fair wage, then, is a weighted average of the wages in the reference group and of the market-clearing wage (1990, 271). Given the potential variation in the determination of the fair wage, we hypothesize the following (in accordance with Fehr et al., (1998a, 1998b, p. 334)):

[3] Again, to reflect the lack of perfect fit between the background theoretical model and the experimental design, we have purposefully made our hypotheses vague and directional rather than precise. Noting that a laboratory model will not perfectly fit a real-world situation, we find the directional results to be of value in terms of understanding differential impacts of policies.

H5. Effort levels will increase with an increase in the wage offered until a cut-off point is reached.
H6. Effort levels in the control treatment will be higher than in Jokipalo (2019).

The first four hypotheses rely on the Stiglitz-Shapiro model for NSC wages being correct as well as the participants attempting to offer NSC wages. If such behavior is assumed to be portrayed by at least a portion of the population (or here, a portion of the sample), the hypotheses should be supported. Judging from other experimental endeavors, effort levels are likely to increase with wages (including but not limited to Fehr, Kirchsteiger, & Riedl, 1993, 1998a, 1998b and Jokipalo 2019). The cut-off point where effort no longer increases may not be observable with an experiment where wages are determined by the employers; after all, it would not be rational for employers to offer more than what is required to induce maximal effort. Even if some employers chose to do so, there would likely not be enough datapoints to deduce a relationship between wage and effort at very high wage levels.

4 Experimental Design

A total of seven experimental sessions with a total of 117 participants were run at the Experimental Social Science Laboratory (ESSL) at University of California, Irvine in February 2020. The experiments were programmed and conducted with z-Tree (Fischbacher, 2007). The experimental design was based on that of Jokipalo (2019), which in turn builds on the original design by Fehr et al. (1993). Instructions for the control session are included as Appendix A. Our design makes some important changes to the design used in Jokipalo (2019) in order to introduce competition for jobs. First, in a session with 20 participants, 12 were assigned as employees and 8 as employers, making the 'unemployment rate' 33% at minimum. Second, employers and employees were not paired randomly by the experimenter and pairings were not fixed; instead, pairs were decided using a job auction and a preference list of potential employers and employees made by the participants in each round. The real effort/real leisure design used in Jokipalo (2019) is maintained. The repeated game format was also kept, albeit altered. In our design, players were not required to interact with the same person multiple times, but they could recognize each other based on screen names.

Experimental procedures

An experimental session lasted approximately 75 min, with an extra 30 min reserved for instructions, an end questionnaire, and payments. After participants had taken their seats, the instructions for the game were read out loud by an experimenter. Furthermore, instructions were printed on paper and handed out to each participant for reference during the experiment. After any remaining questions had been answered, participants were randomly assigned roles in the game and quizzed on their understanding of the payoff formulae. Roles were fixed for the entire experiment.

Before the experiment was officially begun, participants could ask more questions about the functions and tested the real effort task.

Each participant was also given a screen name to signal whether the player was an employee or an employer and included a randomly assigned color to distinguish them from other players. For example, a player could be called *Employee Yellow,* or *Employer Maroon.* Colors with strong common connotations, such as red (female or republican), blue (male or democrat), white (ethnicity or pure or dull), and black (ethnicity or negativity) were avoided.

A total of six identical rounds were played in each session but one.[4] At the beginning of each round, employers made a wage offer in an open auction, meaning that all offers were visible to all players. However, instead of any employee being able to take up any of these offers, employers could pick a maximum of six employees (out of twelve total in a session of 20 participants) who were eligible to accept their offer. Employees could accept any of the offers they were made eligible for by employers. Both employers and employees also ranked each other by preference, and these preferences were used to pair the players for the round. The same rank could be given to several players, in which case their ranks were randomly assigned. If an employee had not been made eligible to accept an employer's offer, or if an employee had not accepted the offer of an employer, they could not be paired together.

As in Jokipalo (2019), employees who were paired with an employer could then complete up to 10 real effort tasks over a four-minute working period. During this time, all participants were also allowed to browse the internet as they wished, given that they did not make any noise or otherwise disturb others. This constitutes 'real leisure' and is the option many people employed today have to perform their work (Corgnet et al., 2015). After the working period had passed, participants were given information on their points for the round and their history in the game before the next round began.

For the control setting, payoff functions are

$$\pi = \mathrm{v} * 0.1\mathrm{e} - \mathrm{w}$$

For the employers, and

$$\pi = \mathrm{w} - \mathrm{c} - \mathrm{E}$$

For the employees. These functions give the payoffs in a situation in which a labor contract is formed, that is, an offer is made and accepted, and the minimum requirement of one correctly completed task is met by the employee. Unmatched employees and employers (or those in a situation in which the minimum requirement is not fulfilled) receive zero points for the round. In the functions, v represents the maximum

[4] Due to technical difficulties, one session only had five rounds. This session had a false start after which the experiment was restarted. Some participants changed roles at this point. Upon inspection, it seems this particular treatment (BI) showed very little variation over time in both wages and effort, so there should be no bias from the session not having data for the sixth round. Therefore, all six rounds are included in the analysis for other treatments.

value (120 experimental currency units—ECU) produced through labor; e is the number of correctly completed tasks; w is the wage (minimum 0, maximum 120 ECU); c is the baseline cost of work (20 ECU); and E is the total number of attempts to solve a task.

Expressed in words, both employers' and employees' profits are impacted by the wage as well as the effort levels chosen, in a manner that is consistent with previous applications of the gift exchange experiment. Employers reap the value produced through labor and pay the promised wage. Employees earn points through wages. Each attempt to solve a task costs the employee one ECU except for the first attempt of each round, which is free of monetary cost. The baseline cost of work, which is set at 20 ECU, represents costs such as commuting or acquiring work-appropriate clothing. The payoff functions are adapted from Fehr et al. (1993). To aid in decisions, participants were provided with a calculator that produced a payoff table for a given wage offer and all possible effort choices.

In addition to these profits, participants received $7 as a show-up fee. Furthermore, employers started off with an endowment of 150 ECU, whereas employees started with nothing. The reasoning behind this was that employers are much more likely to make losses during the game. It is also realistic that a firm starts off with a positive bank balance, whereas private persons often live paycheck to paycheck, or have little cash at hand. If a participant finished the session with zero points or less, they only received the $7 show-up fee, meaning that negative points were ignored at this stage.

Treatments

Two treatments were applied in addition to control sessions. Only one treatment was applied at each experimental session, that is, all participants are only subjected to one treatment (between-subjects design). In the control setting, there were no social security policies in place, but competition was induced in comparison to Jokipalo 2019. The first treatment was an Unemployment Benefit treatment, where *involuntarily* unpaired employees received 15 ECU for the round instead of zero. The results are to be used to compare the Basic Income treatment against the *status quo* in many modern societies. In the second treatment, or the Basic Income treatment, all employees received 15 points in every round on top of what they would have earned in the control setting. (Employers did not receive such benefits, as they were meant to represent firms, not natural persons.)[5] Instructions for treatments were kept identical whenever possible.

End Questionnaire

At the end of the session, participants are asked to fill out a questionnaire before receiving payment. The questionnaire is attached as Appendix B. The rationale for collecting this information on participants is twofold: first, it enables an inquiry

[5] The decision to give BI just to employees instead of to all subjects was made to try to keep the experiment as close as possible to reality. The decision was also made, because employers in the study didn't have the option to refrain from their task. Further. employers in the experiments had bigger incomes.

into potential differences between groups, and especially any potential difference in reactions to a BI policy, later on. Second, if any differences in the samples for different treatments should arise, it would be possible to control for relevant variables in the analysis stage.

Research Ethics

The study was pre-assessed and cleared by the Ethics Committee of the Tampere Region (decision number 73/2019). Participants gave informed consent to the degree possible—they were not given information about the goals or exact subject of the research in order to avoid contaminating the results.

5 Results and Analysis

5.1 Description of the Dataset

After excluding the tainted data from session 3,[6] the dataset consists of data from 100 participants. 40% of participants in all sessions were assigned the role of an employer and six rounds were played in each session except for session 5, which had 5 rounds. There were five occasions of an employer deciding not to make a wage offer at all. As a result, there are 227 observations of wage offers made by employers in the final dataset. Out of these 227 offers, 202 resulted in the employer being paired with an employee.

Table 1 below presents some of the results obtained from the end questionnaire given to participants. Subjects were on average just over 20 years of age and a little over half of them were female. 13% of all participants were students of business or economics; having such a background can influence how participants approach the game. The vast majority, 84%, agreed that they had been given sufficient instructions for the experiment. About half of the participants thought the real effort task was enjoyable. Still, only one in nine subjects (11%) thought that the employers should profit more than the employees in the game. Two thirds (65%) self-reportedly preferred an even split of profits, and the rest (24%) said they thought the employees should make more than the employers. The control treatment has the strongest preference for employers' profits, whereas the BI treatment has the largest share of subjects preferring an equal split. As the end questionnaire was completed after the experiment to avoid tainting results, it is unclear whether this difference is a result of the treatments or sampling error.

[6] See footnote 4.

Table 1 Descriptive statistics of participants' characteristics by treatment

	Total	Control	UB	BI
Number of subjects	**100**	40	20	40
Age (mean)	**20.3**	20.25	20	20.5
Gender (% female)	**58%**	50%	55%	68%
Econ and business students, %	**13%**	20%	5%	10%
% somewhat or strongly agreeing that instructions were sufficient	**84%**	73%	95%	90%
% somewhat or strongly agreeing that task was enjoyable	**51%**	48%	60%	50%
% with a preference for points going to employer/equally/employee	**11/65/24**	18/55/28	5/65/30	8/75/18

5.2 Wage Offers and Effort Levels

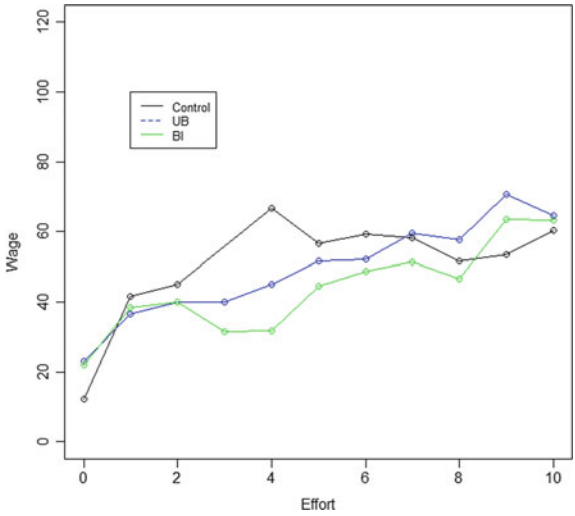
Table 2 below shows the average offers made and the effort exerted in each treatment. In addition, average 'fruitful offers' are shown; an offer was considered fruitful if it resulted in the employer being paired with an employee and the employee exerting at least minimum effort, that is, completing at least one real effort task. The last column shows how many ECU in wages the employees received in exchange for one task completed; since effort decisions were made after determining the wage, this can be construed as the payment required to incentivize employees to complete one task.

It can be observed that offers were clearly highest in the UB treatment (52.62 ECU), with the control and BI treatment showing similar offer levels (47.62 and 45.19 ECU respectively). However, when only fruitful offers are considered, the UB treatment and control show similar levels (55.93 and 54.92 ECU respectively), with the BI treatment falling behind (48.51 ECU). Average effort exerted by those employees who completed at least one task was highest in the control; the wage required to incentivize task completion was also lowest in the control setting. The UB treatment had higher levels of effort exerted compared to the BI treatment, but it seems the wage incentive required was higher.

Graph 1 below gives a more fine-grained look at this relationship, showing the average wage for each possible effort level. This can be construed as the wage level required to induce a particular level of effort. It seems that in the BI and UB setting,

Table 2 Average wage offers and effort levels by treatment

	Offers made	Fruitful offers	Effort exerted	Wage/effort
Control	47.62	54.92	7.05	7.79
Unemployment benefit	52.62	55.93	6.71	8.34
Basic income	45.19	48.51	6.01	8.07
Combined total	**44.14**	**52.64**	**6.59**	**7.99**

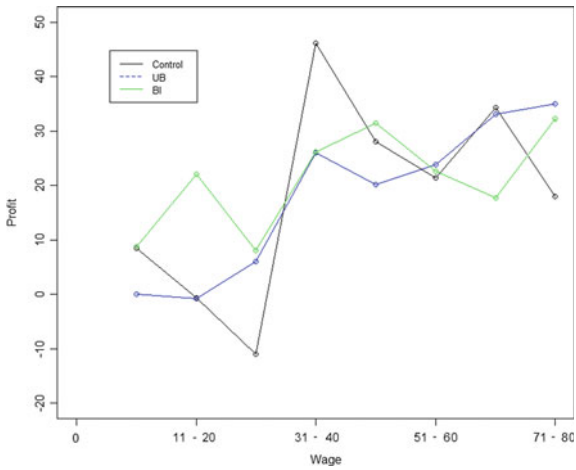


Graph 1 Average wages by effort level

a similar wage offer—about 40 ECU—is required to induce an effort level of 1 or 2 real effort tasks completed. Beyond this point, the UB treatment has higher wage levels associated with each effort level compared to the BI treatment. The control setting has higher wages associated with lower effort levels but is surpassed by both BI and UB at higher effort levels. (Note that there is a missing datapoint at effort level 3 for the control treatment, where coincidentally no subjects chose this effort level at any point.)

Graph 2 shows how much employers profited, on average, according to their wage offer. All offers, not only fruitful ones, are included. The control treatment clearly peaked out at the level where employees just deemed the offer acceptable, so around or just under 40 ECU. A second peak occurred at 61–70 ECU. Paying more than that did not benefit employers even though making profits would have still been feasible in terms of the payoff functions. For the UB treatment, there was also a small peak at around 40 ECU or just lower. A second, higher peak occurred around 70 ECU or higher. Paying any more could not have resulted in higher profits due to the payoff functions. In the BI treatment, there is likewise a peak around 40 ECU or just higher. Again, there is a second peak at 71–80 ECU. However, the BI treatment was also unique in that there was a third peak in profitability right around the point where the employee could break even from working, that is, when their wages just covered their costs of working.[7]

[7] Given that employees in the experiment received BI payments regardless of effort or action, it is conceivable that subjects might have seen them as similar to a larger show-up fee. We are grateful to an anonymous reviewer for raising this potential. There is mixed evidence regarding impacts of differential show-up fees, even within the same type of experiment, (see, for example, Anderson, Mellor & Milyo 2004) so we were not concerned about this potential while designing the experiment. More importantly, show-up fees are generally given at the beginning of an experiment, whereas BI



Graph 2 Average profit by offer level

Regression analysis

Lastly, the data was subjected to regression analysis. Regressions were run separately for effort levels and wage offers. In both cases, multi-level modeling[8] was used to account for (1) individual differences, or personal "baselines" for effort and offers as well as individual reactions to wages offered and effort exerted by the person they were paired with; and (2) any potential differences or clustering between different sessions. In the case of effort, only the individual baselines turned out to be a salient addition to the model, and the results are reported with only that component. For wage offers, session clustering showed no salience, but individual reactions to effort levels were kept in the models. These models are called mixed effects models in the regression tables; where mixed effects are indicated to not be employed, a standard OLS was used. Table 3 reports the results of the effort.

The naïve model (1) shows what is already visible from a crude comparison of means; that the BI treatment has lower effort levels exerted (about one task less per round). At this level, the difference in means is statistically significant. Effort is also lower in the UB treatment, but this difference is not significant. When a mixed effects model is applied in model (2)—that is, interpersonal differences are accounted for—both treatment variables become insignificant, although they stay negative. Beyond this point, there were no considerable differences between models with and without mixed effects, so only the latter are reported.

When the wage level is added as a control in model (3), treatment variables remain insignificant, but the wage is a highly significant determinant of effort; the beta estimate of 0.04 translates to 0.4 more tasks completed per a 10 ECU increase

payments were paid in each round to each subject in the role of employee and can compensate for a wage, should a subject fail to be employed in any given round. This difference is important.

[8] Multi-level regression analyses were run using the 'lme4' package and the 'lmer' function in R Statistics.

Table 3 Regressions on effort levels

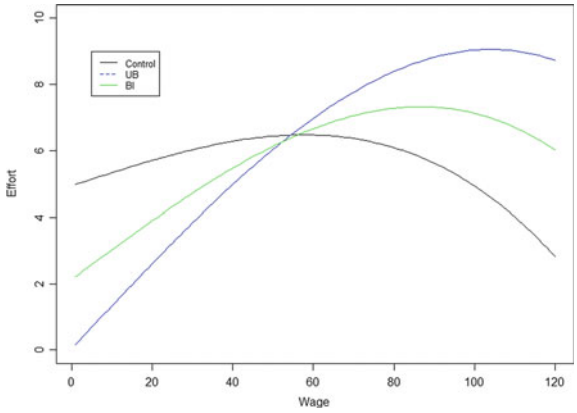
	(1)	(2)	(3)	(4)	(5)
Intercept	**7.05 (0.33)*** **	**6.36 (0.57)*** **	**4.17 (0.75)*** **	**3.26 (0.91)*** **	**4.94 *** **
UB	−0.35 (0.55)	−0.35 (0.97)	−0.24 (0.91)	−0.25 (0.90)	**−4.90 (1.69)** **
BI	**−0.97 (0.47)***	−0.69 (0.82)	−0.45 (0.77)	−0.31 (0.76)	**−2.81 (1.36)***
Offer	–	–	**0.04 (0.01)*** **	**0.07 (0.02)*** **	**0.04 (0.02) ***
Offer^3	–	–	–	−0.000003 (0.000002)	**−0.000004 (0.000002)***
Offer*UB	–	–	**–**	**–**	**0.09 (0.03)** **
Offer*BI	–	–	–	–	**0.05 (0.02)***
Mixed effects	No	Yes	Yes	Yes	Yes
R2	0.01	0.69	0.69	0.69	0.71

*= significant at the 0.05 level. **= significant at the 0.01 level. ***= significant at the 0.001 level

in the wage. Model (4) adds a cubic term for the wage level; this form of relationship allows for effort to rise slowly or remain at zero while wages are very low, then rise rather steeply to level off when wages reach a sufficiently high level. This makes sense intuitively and fits both the data and theory. The cubic term remains just approaching significance in this model. In model (5), all variables are statistically significant. This happens due to the addition of interaction effects between the treatment and wages.

For both treatments, the interpretation of the final model is similar in general: effort is lower at the baseline compared to the control setting. The positive interaction term indicates that participants in the UB and BI treatments also reacted more strongly to the wage they were being paid. This means effort increased more in the treatments as wages increased. Because the UB treatment has lower baseline effort but stronger responses to wage increases, the BI treatment has higher effort in response to wages below approximately 55 ECU. Beyond this point, the UB treatment shows higher effort. The control has the highest effort of the three conditions until wages reach 55 ECU, and the lowest effort when wages are higher than 55 ECU. Graph 3 below visualizes this result.

For wages, a similar analysis is run to gauge the reactions of employers to the treatments; effort exerted by their employee in the previous round; and any potential interaction effects between the treatments and observed effort. The results are depicted in Table 4. As was already mentioned, the mixed effects models for wages include individual baselines for offers and also allow individual employers to react differently to observed effort. All observations are included in model (1), the naïve model without mixed effects. Beyond that, observations from the first round are missing due to the addition of the lagged effort variable. Any cases where an employer was not matched in the previous round are also excluded due to the same reason.



Graph 3 Predicted relationships between wages and effort

Table 4 Regressions on wage offers

	(1)	(2)	(3)	(4)	(5)
Intercept	**47.62 (2.23)*** **	**47.60 (4.62)*** **	**31.92 (4.45)*** **	**16.46 (3.68)*** **	**31.78 (5.41)*** **
UB	5.01 (3.79)	5.03 (7.98)	3.78 (6.42)	5.92 (7.04)	−1.41 (9.80)
BI	−2.42 (3.18)	−2.02 (6.53)	−1.14 (5.19)	**11.0 (5.35)***	1.13 (7.60)
sqrt(lag(effort))	–	–	**7.09 (1.46)*** **	**15.07 (1.53)*** **	**7.17 (2.33)** **
sqrt(lag(effort))*UB	–	–	–	−0.64 (2.93)	2.87 (4.19)
sqrt(lag(effort))*BI	–	–	–	**−5.92 (2.29)***	−1.38 (3.29)
Multi-level model	No	Yes	Yes	No	Yes
R2	0.01	0.67	0.83	0.47	0.83

*= significant at the 0.05 level. **= significant at the 0.01 level. ***= significant at the 0.001 level

As with effort regressions, the relationship between wage offers and effort observed by the employer in the previous round appeared to have a non-linear relationship: wages increase more steeply in response to effort when effort is low, and less when effort is high. In other words, an increase in effort from 2 to 3 tasks resulted in a larger change in wage offers than a change from 8 to 9 tasks, for example. To account for this, the square root of lagged effort is used throughout in place of straightforward lagged effort.

Model (1) again shows a naïve OLS with the information already available: the BI treatment had slightly lower wages and the UB treatment had slightly higher wages on average compared to the control setting. However, these results are not statistically significant. Overall, these results seem to indicate no differences in wage offers

from the treatment itself. Rather, according to the analysis, employers are primarily reacting to the effort levels they observe.

Model (4) is presented here despite it not employing a mixed effects model as it evokes a different interpretation. In model (4), the BI treatment has significantly higher wage offers at the baseline, but a weaker response to effort is observed. These estimates are no longer significant when the mixed effects model is applied; as such, model (5) attributes this difference to individual differences rather than it being an effect of the treatment. As the parameters still maintain the same signs, there is weak evidence for the BI treatment having an effect.

Lastly, the final models for both effort and wages were checked for sensitivity to adding any potential confounders to the equation. Special attention was paid to variables measured in the end questionnaire that were strongly correlated with the treatment and response variables. While there were cases where adding a variable shifted one of the estimates from significant to just approaching significance, the overall interpretations remain unchanged.

6 Discussion

At face value, the UB treatment had the highest wages while the control had the highest effort. The BI treatment had both the lowest wages and the lowest effort, but the UB treatment had a higher "wage to effort ratio" than BI, meaning that employers needed to pay more on average to induce the same effort level. The control had the lowest cost per effort unit for employers. At closer examination, it seems that employees in different settings reacted differently to different wage offers. When wages were low to medium—from 0 to 55 ECU—the highest effort was induced in the control setting and lowest in the UB treatment. Contrarily, at medium to high wages—55 to 120 ECU—effort was highest in the UB treatment and lowest in the control. The BI treatment, then, sits in the middle for all wage levels.

Here it should be noted that there was no taxation in the experimental design; whether a BI transfer that is indirectly taxed away at higher wage levels would have the same effect is unclear. For wages, there seemed to be no difference between the settings when effort exerted by employees was taken into account; only very weak evidence was found for BI potentially increasing wages at the baseline.[9]

Compared to Jokipalo (2019) which had a largely similar design except for the lack of competition in the labor market, average wages were very similar (43.4 in Jokipalo (2019), 44.14 in the present study). It does not seem as though employers felt the need to compete for labor with wage increases, but also did use this position of added power to their advantage. This may also be due to the stark difference in

[9] It should be noted that the unemployment benefit treatment is also free of taxes, and that this design is consistent with previous experiments. Obviously both BI and unemployment insurance are funded by taxes, and the effect of these on behavior in either case is important to understand, but it is first crucial to assess a baseline case of behavior in the presence of UB or BI without taxes. Extending the experiments to include taxation effects is the subject of a future paper.

effort levels: in Jokipalo (2019), the overall average effort was 4.4 tasks per round, whereas it is 6.59 in our study.

In Jokipalo (2019), the BI treatments had higher effort and wages compared to UB and the control at face value. Based on the results from the regression analysis, only the BI treatment with a BI transfer of 20 ECU was found to significantly increase effort. Participants in that treatment also reciprocated wages more (a positive interaction effect). On the other hand, treatment with a BI transfer showed less reciprocity (a negative interaction effect). Conversely, in the study at hand, BI had the lowest wages and effort of all three settings. Regression analysis suggests that BI had lower effort at the baseline but more reciprocity. The same was true for the UB treatment, although the exact parameters are fairly different, leading to very different responses at the same wage level.

There is also some evidence that a BI policy could encourage taking up very low-paying jobs as well. This can be beneficial from the employee's point of view if they find the job itself very rewarding compared to high-paying jobs or if there is a possibility of higher pay later after they accumulate experience. In this sense, the idea that BI may be used as a wage subsidy is supported: employees were willing to work for a wage that just covered the costs of working, presumably because they still had their BI transfer to fall back on. This effect would be presumably stronger in a real-world case.

According to these results, there is no one answer to whether BI would increase or decrease effort/responsible worker behavior. It definitely increases worker power. The outcome is dependent not only on what social policy measures are already in place, but also on the type of jobs under consideration. There appears to be a different effect for low-paying and high-paying jobs (Table 5).

Table 5 Hypotheses and results

Hypothesis	Result
H1. The highest average wage offers will be observed in the UB treatment	Confirmed (although the difference is not statistically significant)
H2. The UB treatment will have the highest offer size required to induce effort levels higher than the minimum	Conflicted—the UB treatment has the highest average wage for effort levels 2–10 pooled, but the control setting has the highest wage for effort level of exactly 2
H3. The lowest average wage offers will be observed in the BI setting	Confirmed (although the difference is not statistically significant)
H4. The control setting will have the lowest offer size required to induce effort levels higher than the minimum	Rejected
H5. Effort levels will increase with an increase in the wage offered until a cut-off point is reached	Confirmed
H6. Effort levels in the control treatment will be higher than in Jokipalo (2019)	Confirmed

Conclusions

Can we implement a social policy that both increases equity and efficiency? If workers have more power, will they behave responsibly? As noted, any potential effects of a given social policy should be explored before implementation. Our results indicate that the impact of a BI policy on wages and effort can be nuanced. Overall, the wages were lower in the BI treatment than in the other treatments. In response to low wage offers, the reciprocity, in terms of effort levels, of subjects in the BI treatment was lower than those in the control group (with no unemployment benefits) but higher than those in the unemployment benefits treatment. At higher wage offers, effort was higher than that in the control group, but lower than in the unemployment benefits group. The BI policy had no effect on labor supply in our experiment. In short, a BI policy has the potential to perform better than many alternatives in both of our goals. BI can combine power with responsibility.

Funding This work was supported by the Finnish Cultural Foundation, Central Finland regional Fund (grant number 30181511) , the Finnish Cultural Foundation, Central Fund (grant number 00200420) and Manufacturing 4.0 - Finnish Strategic Research Council project No. 313396. The sponsors had no role in study design; in the collection, analysis or interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Declaration of Interest Declaration of interest: none.

Appendix 1. Instructions, Control Session

Welcome to the experiment!

We will now read the instructions aloud. You have also received a printed copy of the instructions. The instructions given to each participant are identical. If you would like to ask a question, you can raise your hand at any point.

Roles in the game

Each participant has been randomly assigned as either an *employer* or an *employee.* Out of a maximum of 20 participants in total, eight have been assigned as employers.

Your role will be displayed on your computer screen as the experiment starts and will remain fixed for the entire session. Each participant will also be randomly assigned a fixed screenname.

Anonymity

The game is played anonymously, that is, your identity will not be revealed to other participants at any point in or after the game. Other participants will also receive no other information about you or your decisions unless otherwise stated in the instructions.

Procedures

The experiment consists of several identical rounds consisting of four phases:

Phase 1

Employers **Choose a Wage Offer** Ranging from 0 to 120 Experimental Currency Units.

All offers are visible to all participants.

Employers **Choose Between 0 and 6 Employees** to Whom They Would like to Make a Wage Offer.

Employees only see which offers they themselves have been selected for.

Employers **Rank Their Chosen Employees** in Order of Preference.

The ranking is not revealed to other participants.

Phase 2

Employees **Choose Whether to Accept the Wage Offers** They Have Been Made Eligible for.

These decisions are not revealed to other participants.

Employees **Rank Employers** Whose Wage Offers They Have Accepted in Order of Preference.

The ranking is not revealed to other participants.

Phase 3

Employers and employees are automatically paired using their preference rankings.

The pairings are not visible to other participants.

Employees **who have been paired with an employer can work.**

The working phase will last four minutes, during which employees can solve up to 10 encryption tasks.

All tasks are alike and require no prior knowledge.

The number of tasks completed will only be visible to the employee and her employer. The employer only receives this information after the working phase is completed.

All participants **are free to decide how to use their time** in this phase, given that they do not leave their station or disturb other participants in any way.

For example, you can use a browser on the computer or on your phone. You can access a browser by pressing the Windows button, but DO NOT close this program!

We do not observe how you use your time apart from completing tasks.

Phase 4

Each participant **is shown their points for the round,** as well as the points their employer/employee made in the round, and the number of tasks completed. The next round begins shortly after.

Accumulating points

All employers start the experiment with a starting balance of 150 experimental currency units. ***All employees*** start with a balance of zero.

Points are dependent on the wage level and the number of tasks completed. The choices of employers/employees a participant was not paired with have no effect on her points.

Employers and employees who have not been paired receive zero points for the round. If a paired employee does not correctly complete any encryption tasks in the working phase, she and her employer also receive zero points for the round.

A paired employer **gains** 12 points for each task their employee completes correctly and **loses** the wage she offered. (One experimental currency unit equals one point.)

A paired employee starts with **−20** points due to costs of working. She **gains** the wage offered by her employer. She **loses** the number of points corresponding to the number of tasks she attempts to complete, correct or failed.

$$\text{Points the round for paired employers} \quad = \quad 12 * (\#\text{ of correct tasks}) - \text{wage}$$

$$\text{Points for the round for paired employees} \quad = \quad -20 + \text{wage} - (\#\text{ of tasks}) + 15$$

Please, make sure you understand how the points are calculated before we proceed. Note especially that it is possible to end up with negative points for a round.

If you have zero points or less in total at the end of the game, you will receive no money aside from your $7 show-up fee.

Payment

Your points for all rounds will be added up, converted into dollars, and paid to you in cash after the experiment. For every 12 points you have at the end of the experiment, you will be paid $1 (rounded upwards to the nearest 25 cents). These earnings are added to your $7 show-up fee.

Example 1 If you have 300 points at the end of the experiment, you will be paid $7 + $(300/12) = $7 + $25 = $32.

Example 2 If you have 75 points at the end of the experiment, you will be paid $7 + $(75/12) = $7 + $6.25 = $13.25.

Example 3 If you have zero points or less in total at the end of the experiment, you will be paid $7.

Beginning the Experiment

We are almost ready to begin the experiment. Before the first round, you can get acquainted with the encryption task given to employees.

You can always raise your hand and ask for help after the experiment has started, but if you already have a question in mind, please, ask now.

Appendix 2. End Questionnaire

1.	Age			
2.	Gender	Female	Male	Other/I'd rather not say
3.	Start year of university studies			
4.	Faculty			
5.	Degree program			
6.	Highest attained level of education	High school/secondary		
		Bachelor's degree		
		Master's degree		
		Doctoral degree		
7.	Current work status	Full-time student		
		Part-time		
		Working part-time		
		Working full-time		
8.	Total work experience	Less than one year		
		One to two years		
		Two to three years		
		More than three years		
9.	Your parents' level of education	Father/mother		
		High school/secondary		
		Bachelor's degree		
		Master's degree		
		Doctoral degree		
10.	Population of your home city	Less than 10,000		
		10,001–100,000		
		100,001–1,000,000		
		More than one million		

About the game

11. The instructions provided me with enough information to understand the game.

Strongly disagree 1 2 3 4 5 Strongly agree

12. What additional information would you have needed?
13. I found the encryption task enjoyable.

Strongly disagree 1 2 3 4 5 Strongly agree

14. What was your goal in the game?
15. What was your strategy to achieve your goal?
16. What do you think the game was about?
17. I think it would be most fair if most of the points in the game w to…

…the employer ____ …the employee ____ equally to both ____

Other questions

For the next five questions, choose the point on the scale that best represents your view.

18. 1 = Most people would try to take advantage of me if they got a chance.
    10 = Most people would try to be fair.

1 2 3 4 5 6 7 8 9 10

19. 1 = Government should take more responsibility to ensure that everyone is provided for.
    10 = People should take more responsibility to provide for themselves.

1 2 3 4 5 6 7 8 9 10

20. 1 = Competition is good. It stimulates people to work hard and develop new ideas
10 = Competition is harmful. It brings out the worst in people.

1 2 3 4 5 6 7 8 9 10

21. 1 = In the long run, hard work usually brings a better life.
10 = Hard work doesn't generally bring success—it's more a matter of luck and connections.

1 2 3 4 5 6 7 8 9 10

22. 1 = People can only get rich at the expense of others.
10 = Wealth can grow so there's enough for everyone.

1 2 3 4 5 6 7 8 9 10

23. Choose all suitable answers: do you see yourself as someone who…

A. Is reserved	B. Is generally trusting	C. Tends to be lazy
D. Is relaxed, handles stress well	E. Has few artistic interests	F. Is outgoing, sociable
G. Tends to find fault with others	H. Does a thorough job	I. Gets nervous easily
J. Has an active imagination		

24. Comments and feedback.

References

Akerlof, G. A. (1984). Gift exchange and efficiency-wage theory: Four views. *The American Economic Review, 74*(2), 79–83.

Akerlof, G. A., & Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics, 105*(2), 255–283.

Anderson, L. R., Mellor, J., & Milyo, J. (2004). Social capital and contributions in a public goods experiment. *The American Economic Review Papers and Proceedings, 94*(2), 373–376.

Benndorf, V., Rau, H.A., & Sölch, C. (2018). Minimizing learning behavior in repeated real effort tasks. *Center for European, Governance and Economic Development Research Discussion Papers 343.* University of Göttingen, Department of Economics.

Brandts, J., & Charness G. (2004). "Do labour market conditions affect gift exchange? Some experimental evidence." *The Economic Journal*, 114, 684–708. Royal Economic Society.
Calnitsky, D. (2018). Debating Basic Income. *Catalyst, 1*(23), 63–90.
Calnitsky, D. (2019). Basic income and the pitfalls of randomization. *Contexts, 18*(1), 22–29.
Corgnet, B., Hernán-González, R., & Schniter, E. (2015). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics, 18*(2), 284–301.
Doering, P., & Piore, M. (1971). Internal Labour Markets. *Lexington.*
Fehr, E., Kirchler, E., Weichbold, A., & Gächter, S. (1998a). When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics, 16*(2), 324–351.
Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics, 108*(2), 437–459.
Fehr, E., Kirchsteiger, G., & Riedl, A. (1998b). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review, 42*(1), 1–34.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.
Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.
Füllbrunn, S., Delsen, L., & Vyrastekova, J. (2019). Experimental economics: A test-bed for the unconditional basic income? In Delsen, L. (Ed.) (forthcoming). *Empirical Research on an Unconditional Basic Income in Europe.*
Gilbert, R., Murphy, N. A., Stepka, A., Barrett, M., & Worku, D. (2018). Would a basic income guarantee reduce the motivation to work? An analysis of labor responses in 16 trial programs. *Basic Income Studies*, 13(2). https://doi.org/10.1515/bis-2018-0011.
Haigner, S. Höchtl, W., Jenewein, S., Schneider, F. G., & Wakolbinger, F. (2012). Keep on working: Unconditional basic income in the lab. Basic income studies. *An International Journal of Basic Income Research, 7*(1). https://doi.org/10.1515/1932-0183.1230.
Holler, M. J., Viggo Høst, & Kai Kristensen, (1992) Decisions on strategic markets—An experimental study. *Scandinavian Journal of Management 8*(2), 133–146, ISSN 0956-5221. https://doi.org/10.1016/0956-5221(92)90021-6.
Holler, M. J., & Leroch, M. (2010). Efficiency and justice revisited. *European Journal of Political Economy*, *26*(3), 311–319, ISSN 0176-2680. https://doi.org/10.1016/j.ejpoleco.2009.11.007.
Jokipalo, V. (2019). Basic income, wages, and productivity: A laboratory experiment. *Basic Income Studies*, 14(2).
Kolm, A., & Tonin, M. (2015). Benefits conditional on work and the Nordic model. *Journal of Public Economics, 127*, 115–126. https://doi.org/10.1016/j.jpubeco.2014.05.010.
Le Barbanchon, T., Rathelot, R., & Roulet, A. (2019). Unemployment insurance and reservation wages: Evidence from administrative data. *Journal of Public Economics, 171*, 1–17.
List, J. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy, 114*(1), 1–37.
Marshall, A. (1925). *Principles of economics: An introductory volume* (Repr). Macmillan.
Noguera, J. A., & De Wispelaere, J. (2006). A plea for the use of laboratory experiments in basic income research. *Basic Income Studies*, 1(2).
Shapiro, C. & Stiglitz, J. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review, 74*(3), 433–444.
van der Heijden, E. C. M., Nelissen, J. H. M., Potters, J. J. M., & Verbon, H. A. A. (2001). Simple and complex gift exchange in the laboratory. *Economic Inquiry, 39*(2), 280–297.
Widerquist, K. (2013). *Independence, propertylessness and basic income*. Palgrave Macmillan.
Widerquist, K. (2018). The devil's in the caveats: A brief discussion of the difficulties of basic income experiments. *Cesifo Forum, 19*(3), 30–35.
Yellen, J. L. (1984). Efficiency wage models of unemployment. *The American Economic Review, 74*(2), 200–205.

Voting and Voting Power

Computing the Public Good Index for Weighted Voting Games with Precoalitions Using Dynamic Programming

Jochen Staudacher

1 Introduction

The theory of cooperative games with transferable utility Chakravarty et al. (2015) deals with the outcomes and benefits which players can gain by forming coalitions. A large number of different approaches to distribute the payoff of the grand coalition have been studied in the literature, with the Shapley value Shapley (1953) being the most widely known solution concept. This article focusses on the Public Good index (abbreviated PGI, also known as Holler index or as Holler–Packel index) which was formally proposed by Manfred Holler (1982) and axiomatized by Holler and Packel (1983) as a solution concept for the special class of simple games, i.e. for games in which coalitions are either winning with value 1 or losing with value 0. The PGI assumes that only minimal winning coalitions are relevant for measuring the relative power of players. Holler and Li (1995) propose a non-normalized version of the PGI measuring absolute power which they call Public Value and present as a solution concept for general cooperative games with transferable utility.

Ideological proximity or common economic interests are just two of many reasons why certain coalitions are more likely than others. Hence, transferable utility games with a partition of the player set into disjoint precoalitions (also known as a priori unions) have become an important branch of cooperative game theory, with the generalization of the Shapley value by Owen (1977) being the most widely known solution concept. As pointed out in the book by Owen (1995), pp. 303, the players in such a precoalition have agreed to keep together, but, even though they will do so in most cases, they are not forced to comply. Therefore, the influence of a player needs to be evaluated in a two-stage process. In the external stage, the power of the precoalition is determined, and in the internal stage, the results for the members of the precoalition are computed. In the terminology of cooperative game theory, this two-stage process translates into an external game (also known as the quotient game) between the precoalitions and an internal game within each precoalition. Together with various coauthors Manfred Holler proposed and investigated a total of

J. Staudacher (✉)
Fakultät Informatik, Hochschule Kempten, Bahnhofstr. 61, 87435 Kempten, Germany
e-mail: jochen.staudacher@hs-kempten.de

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_6

six different variations of the PGI with precoalitions in the articles Alonso-Meijide et al. (2010a, b), Holler and Nohn (2009).

This article discusses the efficient computation of the PGI as well as its six variants with precoalitions for weighted voting games, a very important subclass of simple games. Weighted voting games (also known as weighted majority games or weighted games) are an established model for decision-making and voting in committees, panels or boards. There are n players and each player i is allocated a positive weight w_i, which, in some situations, can be interpreted as the number of votes of a voting bloc. A measure or motion gets passed if and only if a certain quota q, normally more than 50 percent of the sum of all weights, is reached or exceeded. Weighted voting games are relevant well beyond classical voting situations in politics, described in Algaba et al. (2007); Kóczy (2021); Kurz (2016). For example, Holler and Rupp employ the PGI and weighted voting games for analysing social networks in a series of recent papers Holler and Rupp (2019, 2020, 2021) whereas the contemporary paper Staudacher et al. (2021) discusses these tools in the context of indirect control power in corporate shareholding structures. These social and economic network applications may involve large numbers of players as well as precoalitions making the need for fast methods for computing the PGI and its variants with precoalitions (expressed in the final paragraph of the paper by Alonso-Meijide et al. (2010b)) a very relevant subject of research.

Power indices for simple games are frequently computed using generating functions, see, e.g. Algaba et al. (2007); Alonso-Meijide et al. (2008); Bilbao et al. (2000), and the paper by Alonso-Meijide and Bowles Alonso-Meijide and Bowles (2005) for the case of precoalitions. If the subsets of players attain only a small number of different weight sums, this method profits Kurz (2016) and fast-access data structures for polynomials with few coefficients in computer algebra systems like Mathematica Tanenbaum (1997) can be employed. In this paper, we use the strongly related, though mathematically less sophisticated, paradigm of dynamic programming for power index computation Staudacher et al. (2021, 2022), Uno (2012). The recent article Staudacher et al. (2021) proposes a new method for computing the PGI for weighted voting games efficiently. Our goal is to extend the algorithm for the PGI to its six variants with precoalitions.

In Sect. 2, we introduce the basic concepts from cooperative game theory, including simple games, the Public Good index (PGI) and its six variants with precoalitions along the lines of Alonso-Meijide et al. (2009, 2010a, b); Holler and Nohn (2009). Section 3 explains how dynamic programming is used to count coalitions efficiently for weighted voting games and discusses the state-of-the-art algorithms for computing the Public Good index from Staudacher et al. (2021) in detail. Section 4 forms the centrepiece of this paper and presents the new algorithms for the six Public Good indices with precoalitions. We point out how our algorithms reflect both the definitions of the indices as well as the different internal division procedures and present a sophisticated new approach for computing the Owen Extended Public Good Index. Section 5 discusses implementations of our new algorithms in C++ including numerical experiments and reports supportive computing times. We end with some concluding remarks and an outlook to open problems in Sect. 6.

2 Preliminaries

In this section, we briefly review some terminology for cooperative games and precoalitions along the lines of the paper by Alonso-Meijide et al. (2009) and define the Public Good index (PGI) and its variants with precoalitions following the articles Alonso-Meijide et al. (2010a, b), Holler and Nohn (2009).

2.1 Cooperative Games, Simple Games and Precoalitions

Let $N = \{1, ..., n\}$ denote a finite set of n players. A group of players $S \subseteq N$ is called a coalition, whereas 2^N symbolizes the set of all subsets of N. $\emptyset$ stands for the *empty coalition* and N is referred to as the *grand coalition*. By $|S|$ we denote the cardinality of a coalition S, i.e. the number of its members, hence $|N| = n$. An n-person cooperative game with transferable utility can be characterized as a pair (N, v) where $v : 2^N \to \mathbb{R}$ is referred to as the *characteristic function* assigning a real value to all coalitions $S \in 2^N$, with $v(\emptyset) = 0$, i.e. in a cooperative game the value of the empty coalition is always zero, see, e.g. one of the books Chakravarty et al. (2015), p. 20, or Owen (1995), p. 213. A cooperative game is *monotone* if for all coalitions $S, T \in 2^N$ the relation $S \subseteq T$ implies $v(S) \leq v(T)$.

We call a cooperative game *simple* if it is monotone and there holds $v(N) = 1$ and $v(S) = 0$ or $v(S) = 1$ for each coalition $S \subset N$. Coalitions for which $v(S) = 1$ are called *winning coalitions* in simple games, whereas coalitions for which $v(S) = 0$ are termed *losing coalitions*. A player i is a *critical player* (also known as a decisive player or swing player) in a winning coalition S if $v(S\backslash\{i\}) = 0$, i.e. the winning coalition S becomes a losing one if player i leaves S. We call a winning coalition S *minimal winning* if it contains only critical players, i.e. if every proper subset of S is a losing coalition.

Weighted voting games (also known as weighted majority games or weighted games) are probably the most important subclass of simple games. They are employed as models in a large number of practical applications Algaba et al. (2007); Holler and Rupp (2019, 2020, 2021); Kóczy (2021); Kurz (2016); Staudacher et al. (2021). An n-player weighted voting game is specified by n non-negative real weights $w_i, i = 1, \ldots, n$, and a non-negative real quota q, normally $q > \frac{1}{2}\sum_{i=1}^{n} w_i$. Its characteristic function $v : 2^N \to \{0, 1\}$ takes the value $v(S) = 1$ for every winning coalition S, i.e. $w(S) = \sum_{i \in S} w_i \geq q$, and $v(S) = 0$ otherwise, implying that coalition S is losing.

Let us create an external division of our set of players $N = \{1, ..., n\}$ into precoalitions (also known as a priori unions). Let $\mathcal{P}(N)$ denote for the set of all partitions of N, with a partition being a set of non-empty subsets of N satisfying the constraint that N is a disjoint union of these subsets. We call an element $P \in \mathcal{P}(N)$ a *coalition structure* (also known as a system of unions) of the set N. A simple game with a coalition structure can be written as a triplet (N, v, P). Following Alonso-Meijide et al. (2009), we write our coalition structure in the form $P = \{P_1, \ldots, P_l\}$,

i.e. we have l precoalitions $P_1, \ldots, P_l$ and the set $L = \{1, \ldots, l\}$ serves as the index set of the partition P. For a weighted voting game with a coalition structure P, the *external game* (also known as the quotient game) is defined as the weighted voting game v^P played between the l precoalitions Alonso-Meijide et al. (2009); Malawski (2004). The external game is formally defined as the weighted voting game $[q; w(P_1), \ldots, w(P_l)]$, i.e. it is characterized by the (unmodified) quota q and the weights $w(P_1), \ldots, w(P_l)$ of the l precoalitions, where $w(P_k) = \sum_{i \in P_k} w_i$ with $k \in L$, $L = \{1, \ldots, l\}$.

2.2 The Public Good Index and Its Variants with Precoalitions

A function f that receives an n-person simple game (N, v) specified by its player set N and its characteristic function v as its input, and passes a unique vector $f(N, v) = (f_1(N, v), \ldots, f_n(N, v))$ as its output is called a *power index*. The literature offers an array of power indices, including the Shapley–Shubik index Shapley and Shubik (1954), the Banzhaf index Banzhaf III (1964), the Johnston index Johnston (1978) and the Deegan–Packel index Deegan and Packel (1978). In this paper, we focus entirely on the Public Good Index (PGI) formally defined by Manfred Holler (1982) and refer the reader to the overview article by Bertini et al. (2013) for a deeper discussion of power indices.

Given a simple game (N, v) with n players, let M denote the set of its minimal winning coalitions and M_i the set of minimal winning coalitions containing player $i \in \{1, \ldots, n\}$. The *Public Good index* (PGI) δ_i of player i is given as

$$\delta_i(N, v) = \frac{|M_i|}{\sum_{j=1}^{n} |M_j|}. \tag{1}$$

A *coalitional power index* g is a function which gets an n-person simple game with a coalition structure (N, v, P) specified by its player set N, its characteristic function v and a partition P as its input and delivers a unique vector $g(N, v, P) = (g_1(N, v, P), \ldots, g_n(N, v, P))$ as its output.

Taking up the idea of the external game v^P played between the precoalitions introduced at the end of Sect. 2.1, we can measure the power of a union $Q \in P = \{P_1, \ldots, P_l\}$ in terms of the PGI. We denote the set of minimal winning coalitions in the external game by M^P and by M_Q^P the set of those minimal winning coalitions in the external game containing precoalition $Q \in P$. Following Alonso-Meijide et al. (2010a), we can write

$$\delta_Q(P, v^P) = \frac{|M_Q^P|}{\sum_k |M_{P_k}^P|}. \tag{2}$$

With the definitions of the PGI on the levels of individual players (1) and precoalitions (2) in place, we can define the six variants of the PGI with precoalitions.

The *Solidarity PGI* Alonso-Meijide et al. (2010b) assigns power to each precoalition according to its PGI in the external game (2) in the first step. In the second step, the Solidarity PGI stresses the public good property and attributes equal power to each member of the same precoalition. We can formally define the Solidarity PGI Υ_i of player $i \in P(i)$, i.e. player i contained in union $P(i)$, as follows:

$$\Upsilon_i(N, v, P) = \delta_{P(i)}(P, v^P)\frac{1}{|P(i)|}. \tag{3}$$

The *Union PGI* Holler and Nohn (2009) reflects the spirit of the original PGI by assuming that the coalitional value is a public good and only minimal winning coalitions (with respect to the coalition structure P) are relevant. The power of an individual player $i \in P(i)$, i.e. player i contained in union $P(i)$, is thus proportional to the number of minimal winning coalitions her precoalition belongs to in the external game. We can formally define the Union PGI Λ_i of player $i \in P(i)$, as follows:

$$\Lambda_i(N, v, P) = \frac{|M^P_{P(i)}|}{\sum_k |P_k||M^P_{P_k}|}. \tag{4}$$

As for the Solidarity PGI, players within the same precoalition obtain the identical Union PGI. We note that among the six extensions of the PGI for precoalitions, the Union PGI is the only solution concept that does not attribute power to precoalitions based on the PGI in the external game (2).

Holler and Nohn (2009) propose three different approaches for reflecting an individual player's threat power to leave the union. In all three cases, the total power attributed to a precoalition Q is given by its external PGI (2). That power is then distributed internally among individual members of unions in threat games.

For the *Threat PGI 1* (TPGI 1) Holler and Nohn (2009), only a minimal degree of stability of the precoalition structure P is assumed. As soon as a single player i leaves her coalition $P(i)$, then not only that union $P(i)$, but the complete precoalition structure P breaks apart. In terms of intra-union allocation of power, this model implies that subsets of a union are not only able to cooperate with other precoalitions, but also with subsets of these precoalitions. Hence, we define the Threat PGI 1 T^1_i of player $i \in P(i)$ as follows:

$$T^1_i(N, v, P) = \delta_{P(i)}(P, v^P)\frac{\delta_i(N, v)}{\sum_{j\in P(i)} \delta_j(N, v)}, \tag{5}$$

whenever $\sum_{j\in P(i)} \delta_j(N, v) > 0$ and $T^1_i(N, v, P) = 0$ otherwise.

The *Threat PGI 2* (TPGI 2) Holler and Nohn (2009) assumes a greater degree of stability of the precoalition structure P. As soon as a single player i leaves her coalition $P(i)$, then only that union $P(i)$ breaks apart into singletons, but the rest

of the precoalition structure remains intact. In terms of intra-union allocation of power, this model implies that subsets of a union are only allowed to cooperate with other precoalitions, but not with subsets of these precoalitions. Following Holler and Nohn (2009), for union $Q \in P$, let $P/Q = P\backslash\{Q\} \cup \{\{i\}|i \in Q\}$ stand for the precoalition structure after Q breaks up into singletons $\{i\}$, $i \in Q$. We define the Threat PGI 2 T_i^2 of player $i \in P(i)$ as follows:

$$T_i^2(N, v, P) = \delta_{P(i)}(P, v^P)\frac{\delta_i(P/P(i), v^{P/P(i)})}{\sum_{j\in P(i)} \delta_j(P/P(i), v^{P/P(i)})}, \tag{6}$$

whenever $\sum_{j\in P(i)} \delta_j(P/P(i), v^{P/P(i)}) > 0$ and $T_i^2(N, v, P) = 0$ otherwise.

The *Threat PGI 3* (TPGI 3) Holler and Nohn (2009) assumes a maximal degree of stability of the precoalition structure P. In case a single player i leaves her coalition $P(i)$, then the rest of that union $P(i)$ remains intact as do all the other precoalitions. Following Holler and Nohn (2009), let $P/i = P\backslash\{P(i)\} \cup \{\{i\}, P(i)\backslash\{\{i\}\}$ stand for the precoalition structure after player i breaks away from her union $P(i)$ and plays on her own. We define the Threat PGI 3 T_i^3 of player $i \in P(i)$ as follows:

$$T_i^3(N, v, P) = \delta_{P(i)}(P, v^P)\frac{\delta_i(P/i, v^{P/i})}{\sum_{j\in P(i)} \delta_j(P/j, v^{P/j})}, \tag{7}$$

whenever $\sum_{j\in P(i)} \delta_j(P/j, v^{P/j}) > 0$ and $T_i^3(N, v, P) = \frac{\delta_{P(i)}(P,v^P)}{|P(i)|}$ otherwise.

The *Owen Extended PGI* Alonso-Meijide et al. (2010a, b), Holler and Nohn (2009), distributes power within precoalitions according to the possibilities which the subsets of this precoalitions possess to form winning coalitions with other precoalitions. We call a subset $S \subseteq Q$ of a precoalition Q an *essential part* with respect to a minimal winning coalition $R \in M_Q^P$ (on the external level) if $S \cup \bigcup_{Q'\in R\backslash\{Q\}} Q'$ is a winning coalition in (N, v) and $T \cup \bigcup_{Q'\in R\backslash\{Q\}} Q'$ is a losing coalition for all true subsets $T \subset S$. We denote the *set of essential parts* with respect to a minimal winning coalition $R \in M^P$ (on the external level) containing player i by $E_i^R(N, v, P)$. The Owen Extended PGI Γ_i of player $i \in P(i)$ is defined as follows:

$$\Gamma_i(N, v, P) = \delta_{P(i)}(P, v^P) \sum_{R\in M_{P(i)}^P} \frac{1}{|M_{P(i)}^P|} \frac{|E_i^R(N, v, P)|}{\sum_{j\in P(i)} |E_j^R(N, v, P)|}. \tag{8}$$

We note that the definition (8) coincides with the "counting PGI" from the work by Malawski (2004) and stress that the Owen Extended PGI manages to be as close as possible in spirit to the extension of the Shapley value to games with a coalition structure proposed by Owen (1977). For the coalition structures $P^n = \{\{1\}, \{2\}, \ldots, \{n\}\}$ and $P^N = \{N\}$, the Owen Extended PGI reduces to the PGI in both cases (just as the Owen value Owen (1977) reduces to the Shapley value in both cases).

3 Dynamic Programming for Computing the Public Good Index

In this section, we introduce the technique of dynamic programming for counting coalitions in weighted voting games efficiently and present the state-of-the-art algorithms for computing the PGI for weighted voting games from the paper Staudacher et al. (2021). Given that Staudacher et al. (2021) is partly a survey paper, the discussion of the PGI in Staudacher et al. (2021) is rather brief and thus the new algorithm merits a more detailed presentation in this section of our article. In other words, we are not computing any PGIs with precoalitions in this section, but prepare the groundwork for doing so in Sect. 4.

3.1 Counting Winning and Losing Coalitions via Dynamic Programming

Every weighted voting game allows for an integer representation Kurz (2016). Therefore, we assume that the weights w_i of the n players in our weighted voting game as well as the quota q are positive integers for the rest of the article. We set $\tilde{w} = w(N) = \sum_{i=1}^{n} w_i$ and assume $q \leq \tilde{w}$. We stress that the algorithms presented in this and the next section are valid for any integer quota with $1 \leq q \leq \tilde{w}$.

Dynamic programming is an algorithmic paradigm based on two pillars. We aim to solve a problem algorithmically by dividing it into subproblems and storing intermediate results efficiently. We employ this paradigm to find out how many subsets $S \subseteq \{1, 2, \ldots, n\}$ there are with weight x, i.e. $w(S) = \sum_{i \in S} w_i = x$, for $x \in \{0, 1, \ldots, \tilde{w}\}$.

Theorem 1 *(see Chakravarty et al. (2015), p. 229). Let $T(i, x)$ be the number of possibilities to write the integer x as a sum of the first i weights $w_1, \ldots, w_i$. For all $i \in \{0, \ldots, n\}$ and all $x \in \{0, 1, \ldots, \tilde{w}\}$, the following recursion delivers $T(i, x)$:*

$$\begin{aligned} T(i,0) &= 1 && \text{for } 0 \leq i \leq n \\ T(0,x) &= 0 && \text{for } x > 0 \\ T(i,x) &= T(i-1,x) + T(i-1,x-w_i) && \text{otherwise.} \end{aligned}$$

Note that the above equations can be interpreted as a boundary condition stating that we can obtain the sum 0 in exactly one way, i.e. via the empty set, as another boundary condition stating that we cannot obtain any sum $x > 0$ without any term and an actual recursion mirroring that the first i weights can deliver a sum $x > 0$ either with or without player i. In practice, T is normally not stored as a two-dimensional table, but efficiently as a vector updated from $T(i-1, x)$ to $T(i, x)$. In our discussion of memory space requirements, we follow the convention by Uno (2012) throughout our paper and omit the need to store the n weights and the corresponding n values of

the power indices. This facilitates a clearer and more concise presentation. According to this convention, Theorem 1 enables us to compute the vector $T(n, x)$ for $x \in \{0, 1, \ldots, \tilde{w}\}$ in $O(n\tilde{w})$ time and $O(\tilde{w})$ memory space. We finally note that it is as simple to update from $T(i-1, x)$ to $T(i, x)$ in Theorem 1 as it is to "downdate" the vector $T(i+1, x)$ to $T(i, x)$ via

$$T(i, x) = T(i+1, x) - T(i, x - w_i). \tag{9}$$

3.2 Computing the Public Good Index via Dynamic Programming

The recent article Staudacher et al. (2021) proposes a new algorithm with a favourable pseudopolynomial complexity for computing the PGI of weighted voting games. We present this algorithm in more detail. In this subsection, we assume the positive integer weights of our n players to be in a descending order, i.e. $w_1 \geq \cdots \geq w_n$.

As pointed out in the textbook Chakravarty et al. (2015), p. 235, it is relatively simple to find the total number $|M|$ of minimal winning coalitions in a weighted voting game via dynamic programming in $O(qn)$ time. We observe

$$|M| = \sum_{i=1}^{n} \sum_{x=q-w_i}^{q-1} T(i-1, x)$$

and note that $\sum_{x=q-w_i}^{q-1} T(i-1, x)$ counts the number of minimal winning coalitions with player i being the player with largest index in the minimal winning coalition.

The recent paper Staudacher et al. (2021) shows that not only $|M|$, but also the cardinalities $|M_i|$ of the sets of minimal winning coalitions containing player i can be found in $O(qn)$ time and $O(q)$ memory space for all players. Following Staudacher et al. (2021), we define the operator $d(S)$ removing the player with largest index from a coalition S. Further, let $w(S) = \sum_{i \in S} w_i$ stand for the weight of coalition S and

$$B(i, x) = |\{S \in 2^N | S \subseteq \{i, \ldots, n\}, i \in S, w(d(S)) < x \leq w(S)\}| \tag{10}$$

for all players $i \in \{1, \ldots, n\}$ and all weights $1 \leq x \leq q$. $B(i, x)$ is the number of coalitions S with i as the player with smallest index such that any coalition S has weight greater or equal x whereas S without its player with largest index has weight less than x. We observe $|M_1| = B(1, q)$. Apart from that observation, Eq. (10) may seem awkward at first, but it helps count $|M_i|$ efficiently without any need to know the player with largest index in any minimal winning coalition contained in M_i. We can obtain B by looping for i from n to 1, as follows.

Theorem 2 *(see Staudacher et al. (2021)) For all* $1 \leq i \leq n$ *and all weights* $1 \leq x \leq q$ *there holds*

$$B(i,x) = \begin{cases} 1 & \text{for } 1 \leq x \leq w_i \\ B(i+1, x-w_i) + B(i+1, x-w_i+w_{i+1}) & \text{for } x > w_i, i < n \\ 0 & \text{otherwise.} \end{cases}$$

Proof By definition of $B(i,x)$ from (10), there holds $B(i,x) = 1$ for $1 \leq x \leq w_i$ as we count the singleton coalition consisting of player i. Furthermore, the statement holds true for $i = n$. In all other cases, i.e. for $x > w_i$ and $i < n$, the recursion means that either player $i+1$ is part of a coalition counted in $B(i,x)$ (first term) or player $i+1$ is not part of a coalition counted in $B(i,x)$ (second term).

$$\begin{aligned} B(i,x) &= |\{S \in 2^N | S \subseteq \{i+1,\ldots,n\}, (i+1) \in S, w(d(S)) < x - w_i \leq w(S)\}| \\ &+ |\{S \in 2^N | S \subseteq \{i+1,\ldots,n\}, (i+1) \notin S, w(d(S)) < x - w_i \leq w(S)\}| \\ &= |\{S \in 2^N | S \subseteq \{i+1,\ldots,n\}, (i+1) \in S, w(d(S)) < x - w_i \leq w(S)\}| \\ &+ |\{S \in 2^N | S \subseteq \{i+1,\ldots,n\}, (i+1) \in S, w(d(S)) < x - w_i + w_{i+1} \leq w(S)\}| \\ &= B(i+1, x-w_i) + B(i+1, x-w_i+w_{i+1}). \end{aligned}$$

Theorem 3 *(see Staudacher et al. (2021)) Let* v *be an* n*-player weighted voting game with positive integer weights sorted in a descending order. With the help of the quantities* $T(i,x)$ *from Theorem 1 and* $B(i,x)$ *from Theorem 2, the Public Good index can be computed in* $O(qn)$ *time and* $O(q)$ *memory space for all players as there holds*

$$|M_i| = \sum_{x=0}^{q-1} T(i-1,x) \cdot B(i, q-x).$$

Proof The statement is true for $i = 1$ as $|M_1| = T(0,0) \cdot B(1,q) = 1 \cdot B(1,q) = B(1,q)$. For $i \geq 2$ we find

$$\begin{aligned} |M_i| &= |\{S \in 2^N | i \in S, w(d(S)) < q \leq w(S)\}| \\ &= |\{S \in 2^N | S = S_1 \cup S_2, S_1 \subseteq \{1,\ldots,i-1\}, S_2 \subseteq \{i,\ldots,n\}, i \in S_2, \\ &\quad w(d(S_2)) < q - w(S_1) \leq w(S_2)\}| \\ &= \sum_{x=0}^{q-1} |\{S_1 \subseteq \{1,\ldots,i-1\} | w(S_1) = x\}| \cdot \\ &\qquad |\{S_2 \subseteq \{i,\ldots,n\} | w(d(S_2)) < q - x \leq w(S_2)\}| \\ &= \sum_{x=0}^{q-1} T(i-1,x) \cdot B(i, q-x). \end{aligned}$$

Since Staudacher et al. (2021) does not list any algorithms, we conclude this section with Algorithm 1 for computing the PGI for a weighted voting game specified by its number of players n, its quota q and its vector w of n weights in $O(qn)$ time and $O(q)$ space.

Algorithm 1 Computing the PGI for weighted voting games

```
1: procedure PGI(n, q, w)
2:     Compute vector T(x) = T(n − 1, x) for x ∈ [0, q − 1] according to Theorem 1
3:     Prepare vector B(x) = B(n, x) for x ∈ [1, q] according to Theorem 2
4:     for i from n to 1 do
5:         |M_i| = Σ_{x=0}^{q−1} T(x) · B(q − x).
6:         if i > 1 then
7:             Update vector B(x) = B(i − 1, x) according to Theorem 2
8:             Downdate vector T(x) = T(i − 2, x) according to Equation (9)
9:         end if
10:    end for
11:    for i from 1 to n do
12:        Compute δ_i = |M_i| / Σ_{j=1}^{n} |M_j|
13:    end for
14:    Return vector δ
15: end procedure
```

4 Computing Public Good Indices with Precoalitions via Dynamic Programming

This section forms the centrepiece of the article. It discusses new dynamic programming algorithms for the six Public Good indices introduced in Sect. 2.2. Thereby, we complement the recent work Staudacher et al. (2022). While Staudacher et al. (2022) generalizes the state-of-the-art algorithms for computing the Banzhaf index Banzhaf III (1964) and the Shapley–Shubik index Shapley (1953); Shapley and Shubik (1954) from the papers by Uno (2012) and Kurz (2016) to the Banzhaf-Owen (1981), Owen (1977) and Symmetric Coalitional Banzhaf Alonso-Meijide and Fiestras-Janeiro (2002) indices via two-level procedures, we hereby extend our algorithm for the PGI from the previous section to its six variants with precoalitions.

As we stated in Sect. 2.1, we assume that there are l precoalitions $P_1, \ldots, P_l$. The *external game* (also known as the quotient game) is defined as the weighted voting game played between the precoalitions Alonso-Meijide et al. (2009); Malawski (2004), i.e. a weighted voting game represented by the (unmodified) quota q and the m weights $w(P_1), \ldots, w(P_l)$ where $w(P_k) = \sum_{i \in P_k} w_i$ with $k \in L$, $L = \{1, \ldots, l\}$. Furthermore, we define p as the maximal size of a precoalition, i.e. $p = \max_{k \in L} |P_k|$, and r as the maximal weight of a precoalition.

In the following, we point out how the ideas and algorithms from the previous section translate into external and internal weighted voting games and algorithms for the six Public Good indices with precoalitions from Sect. 2.2 possessing fairly attractive pseudopolynomial computing times and storage requirements.

Theorem 4 *For a weighted voting game with positive integer weights and a precoalition structure $P = \{P_1, \ldots, P_l\}$, both the Solidarity Public Good and Union Public Good indices can be computed in $O(lq)$ time and $O(q)$ space for all players.*

Proof The proof is trivial. As we can see from the definitions of the Solidarity PGI (3) and the Union PGI (4), the only computational challenge is to find $|M_Q^P|$ for all precoalitions $Q \in P$ on the external level using Algorithm 1. For l unions, this can be achieved in $O(lq)$ computing time and $O(q)$ space. □

Next, we devote one theorem to the three Threat PGIs from Sect. 2.2 each.

Theorem 5 *For a weighted voting game with positive integer weights and a precoalition structure $P = \{P_1, \ldots, P_l\}$ the Threat Public Good indices T^1 can be computed in $O((l+n)q)$ time and $O(q)$ space for all players.*

Proof The definition of the TGPI 1 indices (5) reveals the claim. We need two computations of PGIs using Algorithm 1, one on the level of the l precoalitions, another on the level of the n individual players. This can be achieved in $O((l+n)q)$ computing time and $O(q)$ space. □

For the TPGI 2 indices (6), more PGI computations are needed.

Theorem 6 *For a weighted voting game with positive integer weights and a precoalition structure $P = \{P_1, \ldots, P_l\}$, the Threat Public Good indices T^2 can be computed in $O(l(l+p)q)$ time and $O(q)$ space for all players.*

Proof The definition of the TGPI 2 indices (6) reveals that we need $l+1$ computations of PGIs using Algorithm 1, one on the level of the l precoalitions and another l simulating the break-up of each precoalition into singletons. The number of entities in a PGI computation is bounded by $l+p-1$ with p being the maximal size of a precoalition. According to the conventions of the O-notation, we may state that the computation can be performed in $O(l(l+p)q)$ time and $O(q)$ memory space. □

Computing the TPGI 3 indices (7) means more effort as players break away as singletons leading to one internal game per player.

Theorem 7 *For a weighted voting game with positive integer weights and a precoalition structure $P = \{P_1, \ldots, P_l\}$, the Threat Public Good indices T^3 can be computed in $O(nlq)$ time and $O(q)$ space for all players.*

Proof The definition of the TGPI 3 indices (7) reveals that we need $n+1$ computations of PGIs using Algorithm 1, one on the level of the l precoalitions and another n internal PGIs when each player breaks away individually. The number of entities in a PGI computation in any internal threat game is $l+1$. According to the conventions of the O-notation, we may state that the computation can be performed in $O(nlq)$ time and $O(q)$ memory space. □

Computing the Owen Extended PGI for weighted voting games is more complicated than for the other five PGI variants with precoalitions. From its definition (8), we know that in the internal games we need to work out how many times player $i \in P(i)$ is part of a minimal winning coalition S formed together with other players from $P(i)$ and other precoalitions from the set $P \setminus P(i)$. We formulate a theorem with worst-case estimates for computing the Owen Extended PGI. Its proof describes our algorithm.

Theorem 8 *For a weighted voting game with positive integer weights and a precoalition structure $P = \{P_1, \ldots, P_l\}$, the Owen Extended Public Good indices Γ can be computed in $O(l(l+p)q + lpr^2)$ time and $O(q + pr)$ space for all players.*

Proof We assume the weights of the l precoalitions to be in a descending order, i.e. $w(P_1) \geq \cdots \geq w(P_l)$. The case $r = w(P_1) \geq q$ is fairly simple. There are $\tilde{l} \leq l$ precoalitions with $w(P_k) \geq q$ for all $k = 1, \ldots, \tilde{l}$. We obtain the Owen Extended Public Good indices Γ_i for all players i contained in precoalition P_j for $j = 1, \ldots, l$ by first computing the PGIs for the internal games with the weights $w^{int} = (w_1, \ldots, w_{|P_j|})$ and quota q and then dividing the results by $\tilde{l}$.
For $r = w(P_1) < q$, it is more challenging to compute Γ_i for all players i contained in precoalition P_j for $j = 1, \ldots, l$. Equation (8) tells us to find the relative frequencies with which player $i \in P_j$ is contained in an essential part of a minimal winning coalition $R \in M^P_{P_j}$ on external level. We initialize a vector f with $f(i) = 0$ for $i = 1, \ldots, |P_j|$ for summing these relative frequencies. By $w^{int} = (w_1, \ldots, w_{|P_j|})$ we denote the vector of the individual weights of the members of union P_j and by $w(P_j) = \sum_{i=1}^{|P_j|} w_i^{int}$ its sum. Since $w(\bigcup_{Q' \in R \setminus \{P_j\}} Q') < q$, we start with a preprocessing step and use Algorithm 1 to compute

$$h(i, x) = PGI(|P_j|, x, w^{int}) \tag{11}$$

for all $1 \leq i \leq |P_j|$ and all $1 \leq x \leq w(P_j)$. We first look at minimal winning coalitions $R \in M^P_{P_j}$ on external level such that unions $P_{j+1}, \ldots, P_l$ are not contained in R. Let $T^{ext}(j-1, x)$ for $0 \leq x \leq q-1$ be the vector obtained from Theorem 1 for the weights of the first $j-1$ unions, i.e. from $w(P_1), \ldots, w(P_{j-1})$. We update

$$f(i) = f(i) + \sum_{x=q-w(P_j)}^{q-1} T^{ext}(i-1, x) \cdot h(i, q-x) \tag{12}$$

for all $1 \leq i \leq |P_j|$. Next, we loop for k from $j+1$ to $l-1$ and deal with minimal winning coalitions $R \in M^P_{P_j}$ on external level such that unions $P_{k+1}, \ldots, P_l$ are not contained in R. We update the vector $T^{ext}(j-1, x)$ for $0 \leq x \leq q-1$ successively while omitting the weight $w(P_j)$ of precoalition j, i.e. we compute $T^{ext}_{-j}(k-1, x)$ for $0 \leq x \leq q-1$ using Theorem 1 with the weights $w(P_1), \ldots, w(P_{j-1}), w(P_{j+1}), \ldots, w(P_{k-1})$. We update

$$f(i) = f(i) + \sum_{x=q-w(P_j)-w(P_k)}^{q-w(P_j)-1} T_{-j}^{ext}(k-1,x) \cdot h(i, q-w(P_k)-x) \tag{13}$$

for all $1 \leq i \leq |P_j|$. Finally, with $\delta_{P_j}(P, v^P)$ from (2) and $|M_{P_j}^P|$ we find

$$\Gamma_i(N, v, P) = \delta_{P_j}(P, v^P) \frac{f(i)}{|M_{P_j}^P|}.$$

Computing the values $h(i, x)$ in (11) for all $1 \leq i \leq |P_j|$ and all $1 \leq x \leq w(P_j)$ can be done in at most $O(pr^2)$ time and $O(pr)$ memory space as there are at most $p \geq |P_j|$ members of a precoalition and as the weight of a precoalition is at most $r \geq w(P_j)$. For all l precoalitions, the preprocessing step (11) costs at most $O(lpr^2)$ time and $O(pr)$ memory space. The vector T^{ext} needs additional $O(q)$ memory space justifying the estimate $O(q + pr)$ for storage. As for computing time, we also need to consider the outer loop for $j = 1, \ldots, l$ over all precoalitions as well as the inner loops for $k = j + 1, \ldots, l$ for updating the vector f in (12) and (13). This can be done in $O(l(l + p)q)$ time justifying our total estimate $O(l(l + p)q + lpr^2)$. □

There is one major difference between Algorithm 1 for computing the PGI and our new algorithm for the Owen Extended PGI from Theorem 8. In Algorithm 1 for the PGI, we do not need to know the concrete indices of the players with largest index in a minimal winning coalition (assuming that weights are in a descending order). However, our algorithm for the Owen Extended PGI from Theorem 8 not only assumes that precoalitions are in a descending order by their weights, but it also needs the information on the precoalition with largest index in a minimal winning coalition on the external level. The latter requires an additional loop over precoalitions.

5 Numerical Results and Software

In the recent articles Staudacher et al. (2021, 2022), a powerful software package named EPIC (Efficient Power Index Computation) providing efficient C++ implementations of various power indices (both with and without precoalitions) for weighted voting games was introduced. EPIC is freely available via https://github.com/jhstaudacher/EPIC/. We integrated our new algorithms for the six PGI variants with precoalitions in EPIC. For further details on the internal workings of EPIC, readers are referred to Staudacher et al. (2021, 2022).

For testing our new algorithms, we created a number of games, publicly available at https://github.com/jhstaudacher/EPIC/tree/master/test_cases/precoalitions.

Tables 1 and 2 list computing times of the PGI without precoalitions (PGI), the Solidarity PGI (SPGI), the Union PGI (UPGI), the three Threat PGIs (TPGI1, TPGI2, TPGI3), and the Owen Extended PGI (OPGI) for some of these example problems. The numerical results in Tables 1 and 2 were obtained under Ubuntu 20.04 focal (64

Table 1 Computing times for three test problems with quotas equal to 50 % plus 1 vote

	Problem 1: 741 players, $\tilde{w} = 37064$, 40 precoalitions, max. coal. size 40, avg. coal. size 18, max. prec. weight 2049	Problem 2: 3034 players, $\tilde{w} = 152098$, 60 precoalitions, max. coal. size 78, avg. coal. size 50, max. prec. weight 4150	Problem 3: 3434 players, $\tilde{w} = 72068$, 200 precoalitions, max. coal. size 54, avg. coal. size 17, max. prec. weight 1488
Index	Time (sec)	Time (sec)	Time (sec)
PGI	0.973	95.148	56.413
SPGI	0.42	0.252	0.394
UPGI	0.42	0.252	0.396
TPGI1	1.022	96.985	57.891
TPGI2	2.056	25.887	80.330
TPGI3	26.987	747.713	1274.111
OPGI	38.967	624.871	65.645

bit) on an Intel Core(tm) i7-6600U CPU with a clock speed of 2.60 GHz and 16 GB RAM, i.e. on a standard laptop PC.

In Table 1, we compare three different problems with quotas equal to 50 % plus 1 vote in each case. As reported in Staudacher et al. (2021), the new algorithm for the PGI is very fast, handling problems with more than 3000 players and large quotas in less than two minutes. Not surprisingly, the computing times of the SPGI and the UPGI coincide and are very small since sophisticated computations are performed only on precoalition level. Computing times for TPGI1 are only a little larger than for the PGI confirming our claims from Theorem 5. Comparing computing times for TPGI2, more than three times as many precoalitions (and more players) in problem 3 outweigh the fact that the quota in problem 2 is more than twice the quota in problem 3. As predicted by Theorem 7, the computing times for TPGI3 are much larger than for TPGI2 mirroring the internal games for each individual player. Our new algorithm for OPGI proves to be applicable for large problems and outperforms TPGI3 for test problems 2 and 3. Comparing the computing times of OPGI and TPGI2 for problems 1, 2 and 3 underlines the influence of the preprocessing steps (11) and the maximal weight of a precoalition on our algorithm from Theorem 8.

In Table 2, we study the effects of the quota for another test problem. For PGI, SPGI, UPGI, TPGI1, TPGI2 and TPGI3, these effects on computing times are as predicted by Theorems 3, 4, 5, 6 and 7. For OPGI, the quota has hardly any impact on computing times underlining the fact that the preprocessing steps in (11) are by far the most time-consuming part of the algorithm.

Table 2 Computing times for a test problems with 1031 players, $\tilde{w} = 22031$, 60 precoalitions, max. coal. size 52, avg. coal. size 18 and max. prec. weight 1493 for three different quotas

	$q = 5508(\approx 25\%)$	$q = 11016(\approx 50\%)$	$q = 16524(\approx 75\%)$
Index	Time (sec)	Time (sec)	Time (sec)
PGI	0.447	0.875	1.286
SPGI	0.021	0.037	0.052
UPGI	0.021	0.037	0.052
TPGI1	0.473	0.945	1.330
TPGI2	1.214	2.401	3.536
TPGI3	16.304	32.340	47.582
OPGI	14.482	14.536	14.563

6 Outlook and Conclusions

This article proposes new dynamic programming algorithms for six variants of the PGI with precoalitions Alonso-Meijide et al. (2010a, b); Holler and Nohn (2009) for weighted voting games. All the new algorithms employ an algorithm for computing the PGI efficiently which was recently proposed in Staudacher et al. (2021) and supersedes the previous state-of-the-art approach from Matsui and Matsui (2000). Even though further algorithmic improvements for both the Threat Public Good indices 2 and 3 and the Owen Extended PGI (OPGI) might be possible, we emphasize that our current C++ implementations can be applied for large numbers of players. With the software EPIC Staudacher et al. (2021, 2022) for weighted voting games (with and without precoalitions) and the R package CoopGame Staudacher and Anwander (2021) providing a prototypical implementation of the Public Value by Holler and Li (1995) for cooperative games with transferable utility, there is now publicly available software for many solution concepts proposed by Manfred Holler.

Still, open questions in the context of power index computation abound. Overcoming the limitations of weighted voting games and solving other classes of simple games efficiently is one of them. Wilms (2020) presents dynamic programming algorithms for computing Banzhaf and Shapley–Shubik indices for conjunctions and disjunctions of weighted voting games. It appears rewarding to expand these ideas to other power indices, including the PGI and power indices with precoalitions. Wilms (2020) compares his algorithms with quasi-ordered binary decision diagrams (QOBDDs) Bolus (2011, 2012), i.e. a recent technique based on relational algebra, which has not yet been extended to power indices with precoalitions. These binary decision diagrams provide representations of weighted voting games which come with computational costs for their generation, but are independent of the integer weights and quota. Hence, Wilms (2020) speaks of a minimum quota effect meaning that when using QOBDDs the same bounds for the quota and the sum of all weights can be used as for a minimum sum representation Freixas and Molinero (2009) of the weighted voting game. As we confirmed in Sect. 5, our new algorithms benefit

from small integer weights and quotas. In all cases except for OPGI, a smaller quota implies lower storage requirements.

As stated in our introduction, dynamic programming and generating functions Algaba et al. (2007); Alonso-Meijide and Bowles (2005); Alonso-Meijide et al. (2008); Bilbao et al. (2000) are strongly related. The publicly available doctoral dissertation by Lindner (2004) points out in detail how we can use the recursion from Theorem 1 to find the coefficients of a corresponding generating function and elaborates relations between generating functions and recursions used in dynamic programming algorithms for the Banzhaf and Shapley–Shubik indices. While clearly beyond the scope of this article, studying the mathematical relations between dynamic programming and generating functions for power index computation, including the PGI and its variants with precoalitions, appears to be a very promising field of future research. It bears the potential to lead to even faster algorithms. Furthermore, it could be interesting to study whether a recent result by Koiliaris and Xu (2019) can be used to make the dynamic programming algorithms from Staudacher et al. (2021, 2022) even faster.

Another promising task could be the parallel computation of power indices. Our new algorithms for PGIs with precoalitions presented in the proofs of Theorems 6, 7 and 8 appear to be suitable for parallel processing, given that one can compute internal games independently. We also note that dynamic programming algorithms for the Deegan–Packel index with precoalitions Alonso-Meijide et al. (2011) and the Johnston index with precoalitions Mercik and Ramsey (2017) for weighted voting games have yet to be developed. In terms of practical applications, the author hopes that the new algorithms proposed in this work will prove useful for extending existing studies of social and economic networks Holler and Rupp (2019, 2020, 2021); Staudacher et al. (2021) to precoalitions among the player set and large numbers of players.

Acknowledgements The author thanks his former student Jan Filipp for his advice and assistance in implementing the six new PGIs with precoalitions in C++ and an anonymous reviewer whose comments and suggestions helped improve the paper.

References

Algaba, E., Bilbao, J., & M.,& Fernández-García, J. R. (2007). The distribution of power in the European Constitution. *European Journal of Operational Research,176*(3), 1752–1766. https://doi.org/10.1016/j.ejor.2005.12.002.

Alonso-Meijide, J. M., & Bowles, C. (2005). Generating functions for coalitional power indices: An application to the IMF. *Annals of Operations Research, 137*(1), 21–44. https://doi.org/10.1007/s10479-005-2242-y.

Alonso-Meijide, J. M., & Fiestras-Janeiro, M. G. (2002). Modification of the Banzhaf value for games with a coalition structure. *Annals of Operations Research, 109*(1), 213–227. https://doi.org/10.1023/A:1016356303622.

Alonso-Meijide, J. M., Casas-Mendez, B., Holler, M. J., & Lorenzo-Freire, S. (2008). Computing power indices: Multilinear extensions and new characterizations. *European Journal of Operational Research, 188*(2), 540–554. https://doi.org/10.1016/j.ejor.2007.04.019.

Alonso-Meijide, J. M., Bowles, C., Holler, M. J., & Napel, S. (2009). Monotonicity of power in games with a priori unions. *Theory & Decision, 66*(1), 17–37. https://doi.org/10.1007/s11238-008-9114-2.

Alonso-Meijide, J. M., Casas-Méndez, B., Fiestras-Janeiro, M. G., Holler, M. J., & Nohn, A. (2010a). Axiomatizations of public good indices with a priori unions. *Social Choice and Welfare, 35*(3), 517–533. https://doi.org/10.1007/s00355-010-0451-z.

Alonso-Meijide, J. M., Casas-Méndez, B., Fiestras-Janeiro, M. G., & Holler, M. J. (2010b). Two variations of the Public Good Index for games with a priori unions. *Control & Cybernetics, 39,* 839–855.

Alonso-Meijide, J. M., Casas-Méndez, B., Fiestras-Janeiro, M. G., & Holler, M. J. (2011). The Deegan-Packel index for simple games with a priori unions. *Quality & Quantity, 45*(2), 425–439. https://doi.org/10.1007/s11135-009-9306-z.

Banzhaf, J. F., III. (1964). Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev., 19,* 317–343.

Bertini, C., Freixas, J., Gambarelli, G., & Stach, I. (2013). Comparing power indices. *International Game Theory Review, 15,* 1340004. https://doi.org/10.1142/S0219198913400045.

Bilbao, J. M., Fernández-García, J. F., Jiménez Losada, A., & López, J. J. (2000). Generating functions for computing power indices efficiently. *TOP, 8*(2), 191–213. https://doi.org/10.1007/BF02628555.

Bolus, S. (2011). Power indices of simple games and vector-weighted majority games by means of binary decision diagrams. *European Journal of Operational Research, 210*(2), 258–272. https://doi.org/10.1016/j.ejor.2010.09.020.

Bolus, S. (2012). *A QOBDD-based approach to simple games*. Ph.D. thesis, Christian-Albrechts Universität Kiel, Kiel, Germany. Retrieved May 17, 2022 from https://macau.uni-kiel.de/receive/diss_mods_00009114.

Chakravarty, S. R., Mitra, M., & Sarkar, P. (2015). *A Course on Cooperative Game Theory*. Cambridge University Press.

Deegan, J., & Packel, E. W. (1978). A new index of power for simple n-person games. *International Journal of Game Theory, 7*(2), 113–123.

Freixas, J., & Molinero, X. (2009). On the existence of a minimum integer representation for weighted voting systems. *Annals of Operations Research, 166*(1), 243–260. https://doi.org/10.1007/s10479-008-0422-2.

Holler, M. J., & Rupp, F. (2019). Power in networks: A PGI analysis of Krackhardt's Kite network. In N. Nguyen, R. Kowalczyk, J. Mercik, & A. Motylska-Kuźma (Eds.), *Transactions on Computational Collective Intelligence XXXIV* (Vol. 11890, pp. 21–34). Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-662-60555-4_2.

Holler, M. J., & Rupp, F. (2020). Power in networks and the urban space. In E. Macrì, V. Morea, & M. Trimarchi M. (Eds.), *Cultural Commons and Urban Dynamics* (pp. 37–52). Springer. https://doi.org/10.1007/978-3-030-54418-8_4.

Holler, M. J. (1982). Forming coalitions and measuring voting power. *Political Studies, 30*(2), 262–271.

Holler, M. J., & Li, X. (1995). From public good index to public value: An axiomatic approach and generalization. *Control & Cybernetics, 24,* 257–270.

Holler, M. J., & Nohn, A. (2009). The Public Good Index with threats in a priori unions. *Essays in Honor of Hannu Nurmi, 1,* 393–401.

Holler, M. J., & Packel, E. W. (1983). Power, luck and the right index. *Zeitschrift für Nationalökonomie, 43*(1), 21–29.

Holler, M. J., & Rupp, F. (2021). Power in networks: The Medici. *Homo Oeconomicus, 38,* 1–17. https://doi.org/10.1007/s41412-021-00108-1.

Johnston, R. J. (1978). On the measurement of power: Some reactions to Laver. *Environment & Planning A*, *10*(8), 907–914.

Kóczy, L. Á. (2021). Brexit and power in the council of the European Union. *Games, 12*(2), 51. https://doi.org/10.3390/g12020051.

Koiliaris, K., & Xu, C. (2019). Faster pseudopolynomial time algorithms for subset sum. *ACM Transactions on Algorithms (TALG), 15*(3), 1–20. https://doi.org/10.1145/3329863.

Kurz, S. (2016). Computing the power distribution in the IMF. arXiv Preprint arXiv: 1603.01443.

Lindner, I. (2004). *Power measures in large weighted voting games*. Ph.D. thesis, Universität Hamburg, Hamburg, Germany. Retrieved May 17, 2022 from https://ediss.sub.uni-hamburg.de/bitstream/ediss/715/1/Dissertation.pdf.

Malawski, M. (2004). "Counting" power indices for games with a priori unions. In G. Gambarelli (Ed.) *Essays in cooperative games: Theory and decision library* (Vol. 36, pp. 125–140). Springer. https://doi.org/10.1007/978-1-4020-2936-3_10.

Matsui, T., & Matsui, Y. (2000). A survey of algorithms for calculating power indices of weighted majority games. *Journal of the Operations Research Society of Japan, 2000*(43), 71–86.

Mercik, J., & Ramsey, D.M. (2017). The effect of Brexit on the balance of power in the European Union Council: An approach based on pre-coalitions. In N. Nguyen, R. Kowalczyk, & J. Mercik, (Eds.) *Transactions on Computational Collective Intelligence XXVII* (Vol. 10480, pp. 87–107). Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-319-70647-4_7.

Owen, G. (1977). Values of games with a priori unions. In R. Henn, & O. Moeschlin (Eds.), *Mathematical Economics and Game Theory* (Vol. 141, pp. 76–88). Lecture Notes in Economics and Mathematical Systems. Springer. https://doi.org/10.1007/978-3-642-45494-3_7.

Owen, G. (1981). Modification of the Banzhaf-Coleman index for games with a priori unions. In M. J. Holler (Ed.) *Power, Voting, and Voting Power* (pp. 232–238). Physica. https://doi.org/10.1007/978-3-662-00411-1_17.

Owen, G. (1995). *Game theory* (3rd ed.). Academic Press.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games, 28*(2), 307–317.

Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*(3), 787–792.

Staudacher, J., & Anwander J. (2021). Using the R package CoopGame for the analysis, solution and visualization of cooperative games with transferable utility. Retrieved May 17, 2022 from https://cran.r-project.org/package=CoopGame. R Vignette for package version 0.2.2.

Staudacher, J., Olsson, L., & Stach, I. (2021). Implicit power indices for measuring indirect control in corporate structures. In N. Nguyen, R. Kowalczyk, J. Mercik, & A. Motylska-Kuźma (Eds.), *Transactions on Computational Collective Intelligence XXXVI* (Vol. 13010, pp. 73–93). Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-662-64563-5_4.

Staudacher, J., Kóczy, L. Á., Stach, I., Filipp, J., Kramer, M., Noffke, T., et al. (2021). Computing power indices for weighted voting games via dynamic programming. *Operations Research and Decisions, 31*(2), 123–145. https://doi.org/10.37190/ord210206.

Staudacher, J., Wagner, F., & Filipp, J. (2022). Dynamic programming for computing power indices for weighted voting games with precoalitions. *Games, 13*(1), 6. https://doi.org/10.3390/g13010006.

Tanenbaum, P. (1997). Power in weighted voting games. *Mathematica Journal, 7,* 58–63.

Uno, T. (2012). Efficient computation of power indices for weighted majority games. In K. M. Chao, T. Hsu, & D. T. Lee (Eds.), *International Symposium on Algorithms and Computation. ISAAC 2012* (Vol. 7676, pp. 679–689). Lecture Notes in Computer Science. Springer. https://doi.org/10.1007/978-3-642-35261-4_70.

Wilms, I. (2020). Dynamic programming algorithms for computing power indices in weighted multi-tier games. *Mathematical Social Sciences, 108,* 175–192. https://doi.org/10.1016/j.mathsocsci.2020.06.004.

The Art and Beauty of Voting Power

Sascha Kurz, Alexander Mayer, and Stefan Napel

1 Introduction

Individual voting rights entail a potential to affect collective decisions. Greater numbers of votes controlled by large shareholders, party leaders, delegates to a council or committee, etc. typically increase the respective influence. It is not trivial, though, to tell how much greater the influence of, e.g., a voter wielding 25% of all votes is compared to one wielding only 5%. This holds even when an a priori perspective is adopted, meaning that one purposely leaves aside personal affiliations between the voters and empirical preference information. Various indices try to rigorously quantify voting power in order to address this problem.

For binary collective decision making of the *yes*-or-*no* kind—formalized by simple voting games (cf. von Neumann and Morgenstern 1953, Chap. 10 or Taylor and Zwicker 1999)—prominent examples of voting power indices are the Penrose-Banzhaf, Shapley-Shubik, and Holler-Packel indices (cf. Penrose 1946, Banzhaf 1965, Shapley and Shubik 1954, Holler and Packel 1983). Some of them have been extended to non-binary settings such as the determination of a winner from a set of more than two options by alternative methods of social choice (cf. Kurz et al. 2021, for instance). The respective winner could be a particular law selected from multiple legal drafts, the managing director of the IMF chosen from a shortlist of three nominees, a presidential candidate who is picked from various primary contenders, and so on.

We thank an anonymous referee for his or her thoughtful reading and several helpful suggestions.

S. Kurz
Department of Mathematics, University of Bayreuth, Bayreuth, Germany
e-mail: sascha.kurz@uni-bayreuth.de

A. Mayer · S. Napel (✉)
Department of Economics, University of Bayreuth, Bayreuth, Germany
e-mail: stefan.napel@uni-bayreuth.de

A. Mayer
e-mail: alexander.mayer@uni-bayreuth.de

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_7

Applied to a particular voting body such as a parliament or party convention, the IMF Executive Board, the EU Council, or the US Electoral College,[1] power indices illuminate some of the discrete mathematical structure that underlies collective choices. They identify possible swings between losing and winning coalitions, which make outcomes depend on a voter's behavior, or more generally they measure the potential variation of the winning candidate that derives from a single voter's input to the decision process. Light is cast on either a specific institutional arrangement when the focus is on the influence of distinct members of a voting body relative to another or on multiple competing arrangements. One may evaluate, for instance, the implications of a change of the majority threshold or a switch from plurality voting to a runoff system (e.g., see Maskin and Sen 2016 on plurality voting in US presidential primaries).

So power indices yield insights with political or economic meaning. Common questions are as follows: To what extent can a given shareholder control a corporation and may, perhaps, be held responsible for its actions? Is the voting power of two parties at least approximately proportional to their seat shares in parliament? Is a given allocation of voting rights to delegates from different constituencies (e.g., US states in the Electoral College, member countries in the EU Council, departments in a university senate, etc.) 'fair' under a particular set of normative premises? Etcetera.

This article, however, is *not* pursuing serious questions of any such kind. We here employ a power index for non-binary decisions with a seemingly superficial and primarily visual purpose: we try to convey the hidden beauty of weighted voting and want to exhibit artistic aspects of the power that voters can derive from their voting weights.

The article's main part therefore consists of several pages of color images. Depending on personal taste, they may be of interest and produce enjoyment without any further explanation. At the same time, they represent the result of hours of computer calculations (several weeks, in fact). They give a graphical picture of the formal structure behind collective decision making by three players—individuals or homogeneous groups of voters—on three candidates.

It will be assumed that winners are determined by a plurality vote, an antiplurality vote, or one of the many scoring rules that lie 'in between', such as Borda's voting rule. Other voting methods like the various rules that focus on pairwise majority comparisons (Copeland's rule, Kemeny-Young rule, etc.) are amenable to the same kind of representation. They generate less scintillating results however (cf. Kurz et al. 2020).

Appreciators of art and beauty without interest in the formal framework are welcome to jump to Sect. 4. For all others, we will first provide a short introduction to weighted committee games (Sect. 2).[2] These games generalize traditional weighted

[1] We recommend the contributions in Holler and Nurmi (2013) for a good overview of typical applications of power indices. Somewhat atypical applications are discussed by Kovacic and Zoli (2021) and Napel and Welter (2021). Napel (2019) provides a short introduction to power measurement with many further references.

[2] See Kurz et al. (2020) for details and related literature: the article defines weighted committee games, characterizes and counts equivalence classes for selected voting rules, and provides lists

voting from binary majority decisions to social choice from any finite number of options. We explain the pertinent generalization of the Penrose-Banzhaf power index and how this can literally provide colorful insights into how voting weights determine voters' influence on collective decisions (Sect. 3). Possible economic and political implications are briefly pointed out in Sect. 5.

2 Weighted Committees and Scoring Rules

Binary *weighted voting games* involve a set of n players who, respectively, wield voting weights denoted by $w_1, w_2, \ldots, w_n \geq 0$. The majority threshold or decision quota is set to $q > 0$: the players can jointly pass any proposal that is made to them if the subset of players who support the motion wield a combined voting weight of at least q. For instance, if players 1, 2, and 3 have weights of $w_1 = 40\%$, $w_2 = 35\%$, and $w_3 = 25\%$ and face a quota $q = 51\%$, then at least two players must support a proposal for it to pass. Subsets $S \subseteq N$ of the set N of players with $\sum_{i \in S} w_i \geq q$ are also referred to as winning coalitions, while subsets $T \subseteq N$ with $\sum_{i \in T} w_i < q$ are known as losing coalitions.

Weighted voting games constitute a special kind of (binary) *simple voting games*. The latter do not necessarily require a link between winning or losing to weights and a quota. They merely assume that the full player set N is winning, the empty set $\varnothing$ is losing, and winning is monotonic with respect to set inclusion, i.e., any superset of a winning coalition is also winning.

Simple voting games are commonly specified in set-theoretic terms. This is done either by directly listing a subclass of winning coalitions (typically those that are minimal with respect to set inclusion) or using an indicator function v that takes a set $S \subseteq N$ as its argument and outputs a 1 if and only if S is winning. A simple voting game can, however, also be described as a mapping from the set of all possible profiles of players' preferences over a status quo option a_1 and an alternative motion a_2 that is voted on to the set of possible outcomes, specifying for each preference profile the collective decision a_1 or a_2 that is adopted. *Weighted committee games* follow this route and allow to handle also non-binary decisions.

In particular, the latter consider a finite set N of $n \geq 2$ players and assume that each player $i \in N$ has strict preferences P_i over a set $A = \{a_1, \ldots, a_m\}$ of $m \geq 2$ options that the committee needs to choose from. The set of all $m \cdot (m-1) \cdot \ldots \cdot 1 = m!$ conceivable strict preference orderings on A is denoted by $\mathcal{P}(A)$. Any collective decision rule can then be conceived of as a mathematical mapping $\rho \colon \mathcal{P}(A)^n \to A$. This translates any preference profile $\mathbf{P} = (P_1, \ldots, P_n)$ into a single winning alternative $a^* = \rho(\mathbf{P})$. The respective combination (N, A, ρ) of a set of voters, a

of structurally distinct committees. Mayer and Napel (2021) does similarly for the special case of scoring rules. Kurz et al. (2021) generalizes the Penrose-Banzhaf and Shapely-Shubik indices to committee games. For a practical application of the framework, see Mayer and Napel (2020).

set of alternatives, and a decision rule is referred to as a committee game or as a *committee* for short.

A committee (N, A, ρ) is called a *weighted plurality committee* if the decision rule ρ amounts to each voter i casting $w_i \geq 0$ votes for its favorite option and then selecting the alternative a^* that received the most votes as winner. Similarly, for a *weighted antiplurality committee* the decision rule ρ amounts to each voter i casting $w_i \geq 0$ negative or dissenting votes for its least preferred option and then the alternative a^* that received the fewest dissenting votes becomes the winner. In the case of ties, we suppose that they are resolved lexicographically: if, for instance, $A = \{a_1, a_2, a_3, a_4\}$ and these alternatives, respectively, receive 3, 4, 0, and 4 plurality votes from $n = 11$ voters with a weight of $w_i = 1$ each, then a_2 rather than a_4 is chosen. Declaring both a_2 and a_4 to be winners and tossing a coin to reach a resolute decision would be a possibility too. But randomness would complicate the mathematical exposition without changing the illustrations below.

(Weighted) plurality and antiplurality committees are special cases of (weighted) *scoring committees*. These entail the application of a *scoring rule*: the winning candidate or option a^* always is the one that received the highest total score from the voters. Candidates' scores are determined by their positions in each voter's preference ranking and a given vector $\mathbf{s} = (s_1, s_2, \ldots, s_m)$ with $s_1 \geq s_2 \geq \ldots \geq s_m$ and $s_1 \neq s_m$: when voters' weights are $w_1 = \ldots = w_n = 1$, any alternative $a \in A$ receives s_1 points for every voter that ranks a first, s_2 points for every voter that ranks a second, and so on. When the voters have non-uniform weights $w_1, \ldots, w_n \geq 0$, the respective points derived from how voter $i \in N$ ranks the alternatives are multiplied by w_i.

For illustration, suppose that a committee—perhaps the board of a sports club—involves four voter groups, i.e., players $N = \{1, 2, 3, 4\}$, with group 1 wielding 5 votes, group 2 having 4 votes, group 3 wielding 3 votes, and group 4 having only one vote. The weights are summarized by $\mathbf{w} = (5, 4, 3, 1)$. The voters must select one of three candidates, say, Ann, Bob, or Clara, to lead their club.

Let the players' preferences $\mathbf{P} = (P_1, P_2, P_3, P_4)$ rank the candidates as in the following table:

	P_1	P_2	P_3	P_4
1st best	Bob	Ann	Ann	Bob
2nd best	Clara	Clara	Bob	Clara
3rd best	Ann	Bob	Clara	Ann

Using the scoring vector $\mathbf{s} = (1, 0, 0)$ amounts to a weighted plurality vote: Ann receives a total score of $5 \cdot 0 + 4 \cdot 1 + 3 \cdot 1 + 1 \cdot 0 = 7$; Bob's score is $5 \cdot 1 + 4 \cdot 0 + 3 \cdot 0 + 1 \cdot 1 = 6$; and Clara, being ranked first by nobody, gets a score of 0. Ann wins.

Had the above committee used the scoring vector $\mathbf{s} = (1, 1, 0)$ instead, Clara would have won with a score of 10 vs. 7 for Ann and 9 for Bob. The latter vector $\mathbf{s}$ is equivalent to conducting an antiplurality vote because minimizing the number

of dissenting votes is the same as maximizing the number of non-dissenting votes captured by $\mathbf{s} = (1, 1, 0)$.

An example of a voting rule in between plurality and antiplurality is *Borda's rule*: voters state their full preferences and each candidate a receives as many points from a given voter i as there are candidates that i ranks below a. For instance, Bob would receive 2 points for each vote wielded by group 1, 0 points from group 2, 1 point for each of the votes held by group 3, and again 2 points from group 4. This gives Bob a total Borda score of $5 \cdot 2 + 4 \cdot 0 + 3 \cdot 1 + 1 \cdot 2 = 15$. That number is greater than the analogous figures of 14 for Ann and 10 for Clara. So Bob would win if scoring vector $\mathbf{s} = (2, 1, 0)$ or Borda's rule were used.

Maximizing the total score given the scoring vector $\mathbf{s} = (2, 1, 0)$ is equivalent to maximizing the total score for vectors $\mathbf{s}' = (1, 1/2, 0)$ or $\mathbf{s}'' = (4, 3, 2)$. Vector $\mathbf{s}'$ merely halves above numbers, while preserving the order of Ann's, Bob's, and Clara's totals. Similarly, using $\mathbf{s}''$ raises all candidates' scores by $(5 + 4 + 3 + 1) \cdot 2 = 26$ without changing their order. In particular, scoring winners are invariant to positive affine transformations of the adopted scoring vector $\mathbf{s}$. Hence, whenever a committee picks a winner from three candidates by a scoring rule—plurality, antiplurality, Borda, or any other rule that determines the winner by evaluating the candidates' positions in the applicable preference profile $\mathbf{P}$ with decreasing scores—it is without loss of generality to suppose a vector $\mathbf{s} = (1, s, 0)$ such that $0 \leq s \leq 1$.

When a committee with player set N and voting weights $\mathbf{w} = (w_1, \ldots, w_n)$ decides on a set A of $m = 3$ alternatives and uses a decision rule ρ that amounts to applying the scoring vector $\mathbf{s} = (1, s, 0)$, we write $(N, A, r^s|\mathbf{w})$ instead of (N, A, ρ). We refer to such committee as a (weighted) *s-scoring committee* (see Mayer and Napel 2021).

We have seen that the special s-scoring committees with $s = 1$, $s = 1/2$, and $s = 0$ amount to weighted plurality, Borda, and antiplurality committees. As the above example illustrates, the respective committees differ for the considered voting weights $\mathbf{w} = (5, 4, 3, 1)$. Namely, they select a different winner from three candidates for at least some configuration of preferences. Similarly, two plurality committees ($s = 0$) are different depending on whether weights $\mathbf{w} = (5, 4, 3, 1)$ or weights $\mathbf{w}' = (5, 1, 1, 1)$ apply to the players (club members, shareholders, parties, etc.): for the profile $\mathbf{P}$ at hand, Bob rather than Ann would be selected if $\mathbf{w}$ were replaced by $\mathbf{w}'$.

We call two committees (N, A, ρ) and (N, A, ρ') that never select different winners from set A—no matter which preference profile $\mathbf{P} = (P_1, \ldots, P_n)$ is considered—*equivalent*. This means that the respective mappings $\rho\colon \mathcal{P}(A)^n \to A$ and $\rho'\colon \mathcal{P}(A)^n \to A$ are identical, denoted by $\rho \equiv \rho'$. We can have $\rho \equiv \rho'$ even though the verbal descriptions of ρ and ρ' may differ. For instance, ρ may be described as plurality voting with weights $\mathbf{w}' = (5, 1, 1, 1)$ and ρ' as the dictatorship of voter 1: the committee in either case always chooses the alternative that is ranked first according to P_1.

When two s-scoring committees $(N, A, r^s|\mathbf{w})$ and $(N, A, r^s|\mathbf{w}')$ with $\mathbf{w} \neq \mathbf{w}'$ are equivalent, i.e., $r^s|\mathbf{w} \equiv r^s|\mathbf{w}'$, we learn that it does not matter which of the two voting weight distributions prevails: decisions will coincide. From the perspective

of an outsider who does not care about the labeling of the players, this is also true if weights $\mathbf{w}''$ are used that only label the players differently than $\mathbf{w}$.

Consider, for instance, $\mathbf{w}'' = (1, 3, 4, 5)$ instead of $\mathbf{w} = (5, 4, 3, 1)$ in our example. This represents the same abstract decision structure except that player numbers have changed. In particular, the situation for the preferences $\mathbf{P} = (P_1, P_2, P_3, P_4)$ depicted in the table above for weights $\mathbf{w}$ (with Ann winning under plurality rule, Clara under antiplurality rule, etc.) is the same as that with weights $\mathbf{w}''$ and preferences $\mathbf{P}'' = (P_4, P_3, P_2, P_1)$. We then say that $\mathbf{w}$ and $\mathbf{w}''$ are *structurally equivalent* under the considered s-scoring rule: the implied mappings $r^s|\mathbf{w}$ and $r^s|\mathbf{w}''$ become equivalent after suitably relabeling the players.

Having fixed a scoring rule, such as r^s for $s = 1$, the set of all weights $\mathbf{w} = (w_1, \ldots, w_n)$ that are structurally equivalent to a given reference distribution of weights $\tilde{\mathbf{w}} = (\tilde{w}_1, \ldots, \tilde{w}_n)$ can be grouped together and form an *equivalence class* of weights: if two weight distributions $\mathbf{w} \neq \mathbf{w}'$ belong to the same class, the corresponding s-scoring committees always produce identical decisions (once labels of the players are harmonized). If the weight distributions belong to different classes, there exists at least some preference configuration $\mathbf{P}$ that results in different committee decisions.

For antiplurality rule ($s = 1$) and three players ($n = 3$), it turns out that there are only five different equivalence classes—namely those that correspond to reference weights of $\tilde{\mathbf{w}} = (1, 0, 0)$, $(1, 1, 0)$, $(1, 1, 1)$, $(2, 1, 1)$, and $(2, 2, 1)$. Any other distribution of weights among three players is structurally equivalent to one of these, i.e., leads to the same decisions after suitable relabeling (cf. Kurz et al. 2020). Similarly, there are only six structurally different plurality committees for three players. The respective reference weights equal the five just listed for antiplurality rule in addition to $\tilde{\mathbf{w}} = (3, 2, 2)$.

The numbers of structurally distinct s-scoring committees for $s = 1/2$ (Borda) and, more pronouncedly, for $0 < s < 1/2$ or $1/2 < s < 1$ are much higher than those for $s = 0$ and $s = 1$. Exact values have not been published for all s yet, but Mayer and Napel (2021) provide the numbers of equivalence classes for all s that are integer multiples of $1/20$. These numbers range up to 229 and exhibit an M-shaped pattern reproduced in Fig. 1.

Knowing that a given weight distribution among three players structurally amounts to, say, $(2, 1, 1)$ can simplify the analysis of the respective committee: the distribution of voting power is as if weights were $(2, 1, 1)$. So are players' manipulation incentives, strategic voting equilibria, the scope for voting paradoxes, etc.

Alas, it is generally an arduous task to determine for a given weight distribution $\mathbf{w}$ to which scoring equivalence class it belongs (for fixed vector $\mathbf{s}$). The respective equivalence classes form convex polyhedra that are defined by linear inequalities. When we consider three players and restrict attention to their relative voting weights $\bar{\mathbf{w}} = \mathbf{w}/(w_1 + w_2 + w_3)$ (so that $\bar{w}_1 + \bar{w}_2 + \bar{w}_3 = 100\%$), the polyhedra are either points, lines, or area pieces bounded by lines. They jointly cover the triangle highlighted in Fig. 2 below—the so-called two-dimensional unit simplex.

Suppose that we have a 'map' of all equivalence classes in this simplex. Then one may start out with an arbitrary weight distribution $\mathbf{w} = (w_1, w_2, w_3)$, compute the

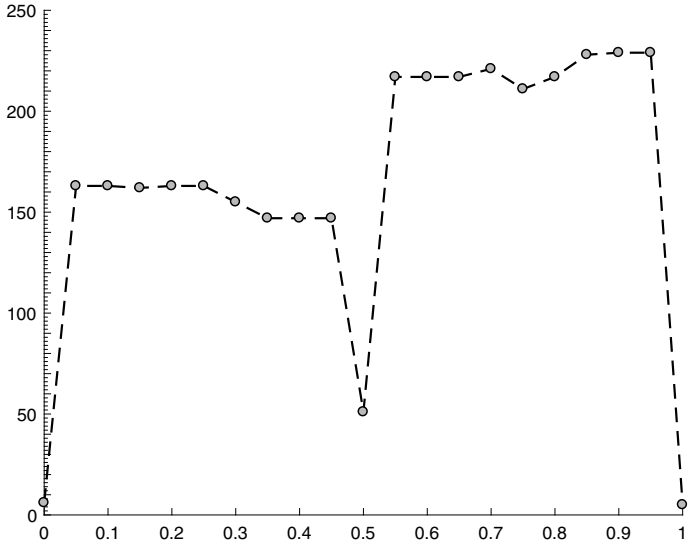


Fig. 1 Number of s-scoring committees for $m = 3$ and $s \in \{0, 0.05, 0.1, \ldots, 1\}$

corresponding relative weight distribution $\bar{\mathbf{w}}$, locate it in the simplex map, and now identify the applicable class.

Such simplex maps can indeed be constructed. Namely, the figures depicted in Sect. 4 show the links from all possible weights to equivalence classes, except that we leave out a legend that would identify the respective equivalence classes via reference distributions of weights.[3]

3 Voting Power and Color

Before we present our figures, let us explain how the selected coloring relates to the a priori voting power of the three players involved. An index of voting power generally takes a description of a voting body—a simple voting game or, in our case, an s-scoring committee of three players deciding on three options—as its input and produces a real number for each player as its output. The respective numbers reflect the players' influence on collective decisions according to a specific conception of a

[3] See Fig. 5 in Mayer and Napel (2021). It provides a map of the 51 Borda equivalence classes for $n = m = 3$ and $w_1 \geq w_2 \geq w_3$ with a reference distribution of weights for each class. Maps could be constructed for more than three alternatives, too, but the higher number of preference profiles and perturbations has considerable computational costs. Equivalence classes for scores $0 < s < 1$ change fast: the number of Borda classes rises from 51 to 505 and ≥ 2251 for $m = 3, 4$, and 5 (Kurz et al. 2020). Corresponding analogues of Fig. 3 exhibit smoother transitions with even more shades of color.

player being influential. They are based on specific probabilistic assumptions about the voting situations faced by the players.

The popular Penrose-Banzhaf power index (Penrose 1946, Banzhaf 1965) equates 'being influential' with the possibility of the considered player changing or swinging a decision at hand if the preferences and behavior of all other players are held fixed. This possibility arises, for instance, in unweighted binary majority decisions when the other players are split equally into a *yes*-camp and a *no*-camp, so that the player in question can determine which option receives the majority. In other words, the considered player's vote is pivotal for the outcome. The Penrose-Banzhaf index assesses the probability of pivotality events for a given player under the assumption that all other players vote *yes* or *no* with equal probabilities and independently of another. This is equivalent to assuming that all *yes*-or-*no* configurations or all coalitions $S \subseteq N$ of players who support a change of the status quo are equally likely.[4]

When the collective decision requires a choice from three or more options, such as candidates Ann, Bob, and Clara above, one can similarly identify 'being influential' with the committee's decision depending on or varying in the considered player's preferences. For instance, if in our sports club example, player 3 did not rank Ann before Bob and Clara but had Bob as its first preference before Ann and Clara, then the plurality winner would be Bob rather than Ann. Hence, player 3 is pivotal in the considered voting situation. So is player 2, whereas players 1 and 4 have no scope to individually change the winner for the given preferences of the respective others. Players 1 and 4 are, however, pivotal for many other preference configurations $\mathbf{P} = (P_1, P_2, P_3, P_4) \in \mathcal{P}(A)^4$ that may arise. So also they are influential from an a priori perspective that considers all preference combinations to be possible.

Just as the Penrose-Banzhaf index is based on independent and equiprobable *yes*-or-*no* preferences in the binary case, players' preferences P_i will be assumed to be distributed independently from the others also for more than two options, assigning equal probability to each of the conceivable strict orderings of the options. When assessing the a priori influence implications of voting weights $\mathbf{w} = (5, 4, 3, 1)$, for instance, we will therefore assume player 1 to be as likely to rank (i) Ann before Bob before Clara, as to rank (ii) Ann before Clara before Bob, (iii) Bob before Ann before Clara, (iv) Bob before Clara before Ann, (v) Clara before Ann before Bob, or (vi) Clara before Bob before Ann. We allow the same six possibilities to arise independently also for players 2, 3, and 4. So there are a total of $6 \cdot 6 \cdot 6 \cdot 6 = 6^4 = 1296$ different voting situations that are equally probable when four players decide on three options.

We will focus here on only $n = 3$ players who decide on $m = 3$ options, so that $(m!)^n = 6^3 = 216$ different preference profiles $\mathbf{P} = (P_1, P_2, P_3)$ are possible. Holding a particular player of interest, say player $i \in \{1, 2, 3\}$, fixed, we check for each profile whether a change of i's ranking P_i to one of the alternative five rankings

[4] The Shapley-Shubik index (1954) belongs to the same family of indices but supposes a positive correlation of *yes*-or-*no* preferences across voters. In technical terms, it assumes an *impartial anonymous culture (IAC)*, while the Penrose-Banzhaf index reflects an *impartial culture (IC)*. The Holler-Packel index (1983) does not consider all coalitions of *yes*-supporters but only *minimal winning coalitions* $S \subseteq N$ in which every *yes*-vote is pivotal for the outcome.

P_i' would make a difference to the collective decision. Whenever this is the case, i.e., the profile $\mathbf{P}'$ that is created by replacing P_i in $\mathbf{P}$ by P_i' yields a collective decision $r^s|\mathbf{w}(\mathbf{P}') \neq r^s|\mathbf{w}(\mathbf{P})$, we count this as a *swing position* for player i. Player i's power index value is then taken to be the ratio of actual swing positions to the maximum conceivable number of such positions.

The latter corresponds to the number of swing positions that a dictator player would hold. For each of the 216 possible preference profiles of three voters on three options, the collective choice under a dictatorship equals the dictator's most preferred alternative. So starting from given preferences of the dictator over three candidate, say ranking (i) above, a switch to four of the five alternative rankings produces a different winner—namely preference changes from (i) to (iii), (iv), (v), or (vi). These perturbations involve a different top preference than (i) and let Bob or Claire win instead of Ann. It follows that a dictator player has $216 \cdot 4 = 864$ swing positions: they derive from considering 216 distinct voting situations and, for each situation, checking all five ways to spontaneously change the dictator's ranking of the options. Such a change might reflect an idiosyncratic change of mind, perhaps due to new private information on the candidates; it might arise because the player is corrupt and sells its vote to an outside agent; it could simply be a demonstration of the player's power; etc. If a player i in the actual scoring committee should have 432 swing positions, then the corresponding ratio $432/864 = 1/2$ reveals i to be half as powerful as a dictator would be.

Expressing this reasoning in general mathematical terms leads to the (generalized) Penrose-Banzhaf index

$$\mathcal{PBI}_i(N, A, \rho) = \frac{\sum_{\mathbf{P}\in\mathcal{P}(A)^n} \sum_{P_i'\neq P_i\in\mathcal{P}(A)} \Delta\rho(\mathbf{P}; P_i')}{m!^n \cdot (m! - (m-1)!)} \tag{1}$$

of player i's a priori influence or *voting power* in committee (N, A, ρ), as introduced and axiomatically characterized by Kurz et al. (2021).[5] Here $\Delta\rho(\mathbf{P}; P_i')$ denotes an indicator function that is 1 if $\rho(\mathbf{P}') \neq \rho(\mathbf{P})$, and 0 otherwise. Equation (1) is summing over all voting situations (i.e., all conceivable preference configurations $\mathbf{P}$), counts the number of changes of mind by player i (i.e., perturbations of i's preferences P_i to some $P_i' \neq P_i$) that change the collective decision, and then divides this by the total number of swing positions for a hypothetical dictator player ($6^3 \cdot 4 = 864$ for $m = 3$ options and $n = 3$ players). So for an s-scoring committee $(N, A, r^s|\mathbf{w})$ of three players, the triplet

$$\big(\mathcal{PBI}_1(N, A, r^s|\mathbf{w}), \mathcal{PBI}_2(N, A, r^s|\mathbf{w}), \mathcal{PBI}_3(N, A, r^s|\mathbf{w})\big),$$

[5] Replacing the IC assumption that underlies Eq. (1) by the IAC assumption (cf. fn. 4) naturally generalizes the Shapley-Shubik index (see Kurz et al. 2021). By contrast, generalization of the Holler-Packel index would first require the definition of a suitable analogue of minimal winning coalitions in weighted committee games. One possibility would be to study each winning alternative $a \in A$ separately and to consider *a-minimal preference profiles* $\mathbf{P}$ where $\rho(\mathbf{P}) = a$ such that $\rho(\mathbf{P}') \neq a$ for any profile $\mathbf{P}'$ in which a is ranked lower by some voter with constant preferences on subset $A \setminus a$.

or $\mathcal{PBI}$ for short, quantifies the distribution of voting power in the committee in terms of how close the individual players are to having dictatorial influence. In our sports club example, the power distribution amounts to $\mathcal{PBI} \approx (0.6296, 0.4815, 0.4444, 0.0741)$. That is, player 1 has about 63% of the influence of a dictator while player 4 only has about 7% of the influence of a dictator. The influence of players two and three is just under 50% of that of a dictator.

For graphical purposes, one might now associate player 1's power value $\mathcal{PBI}_1$ with the color red, player 2's power $\mathcal{PBI}_2$ with green, and player 3's power $\mathcal{PBI}_3$ with blue. Thus, we would have linked the scoring rule r^s for a given value of s and a particular distribution $\mathbf{w}$ of voting weights to a particular color using the common RGB color code. For instance, $\mathcal{PBI}(N, A, r^s|(1, 0, 0)) = (1, 0, 0)$ for any $0 \leq s \leq 1$ and this would correspond to bright red color. Or the power distribution $\mathcal{PBI} = (588/864, 516/864, 312/864) \approx (0.6806, 0.5972, 0.3611)$ that is derived by Kurz et al. (2021) for $s = 1/2$ and weights $\mathbf{w} = (6, 5, 3)$ would correspond to a dark khaki color.

Although this would be feasible, the figures in Sect. 4 will not use exactly this coloring option. We will rather make two modifications: first, we will adopt a structural view on committee equivalences, i.e., we do not consider player labels important. Hence, we will give the same color to all six points in the unit simplex that represent relative voting weights of, e.g., $\bar{\mathbf{w}} = (6/14, 5/14, 3/14)$ after sorting the weights in decreasing order. This implies that the coloring of the weight simplex will be threefold radially symmetric around $\bar{\mathbf{w}} = (1/3, 1/3, 1/3)$, as well as mirror symmetric with the three symmetry axes $\bar{w}_1 = \bar{w}_2$, $\bar{w}_2 = \bar{w}_3$, and $\bar{w}_1 = \bar{w}_3$.

Second, we will apply a transformation when turning power triplets $\mathcal{PBI}$ into RGB levels. The motivation is to make better use of the available color palette, to obtain a somewhat lighter image than by, e.g., associating $\bar{\mathbf{w}}$ with dark khaki, and to represent dictatorial power by the dark blue color that has already been used, e.g., by Kurz et al. (2020).

4 Simplex Maps of Equivalence Classes

All images displayed in this section are derived via the following five steps:

1. We fix a scoring vector $\mathbf{s} = (1, s, 0)$ and consider the corresponding scoring rule r^s for collective decisions on $m = 3$ options by $n = 3$ players.
2. We use a finite grid of rational numbers and let the computer loop through all relative voting weight distributions $\bar{\mathbf{w}}$ with $1 \geq \bar{w}_1 \geq \bar{w}_2 \geq \bar{w}_3 \geq 0$ on this grid.
3. For each of the 282 376 weight distributions $\bar{\mathbf{w}}^k$, $k = 1, 2, \ldots, 282\,376$, on the adopted grid, we compute the Penrose-Banzhaf voting power $\mathcal{PBI}^k$ in the respective weighted s-scoring committee $(N, A, r^s|\bar{\mathbf{w}}^k)$.
4. The obtained triplet $(\mathcal{PBI}^k_1, \mathcal{PBI}^k_2, \mathcal{PBI}^k_3)$ is then transformed into red, green, and blue intensities $(R, G, B) = \left(\frac{2\cdot\mathcal{PBI}^k_3}{\max_k \mathcal{PBI}^k_3}, \frac{\mathcal{PBI}^k_2}{\max_k \mathcal{PBI}^k_2}, \frac{\mathcal{PBI}^k_1 - \min_k \mathcal{PBI}^k_1}{1 - \min_k \mathcal{PBI}^k_1}\right)$.

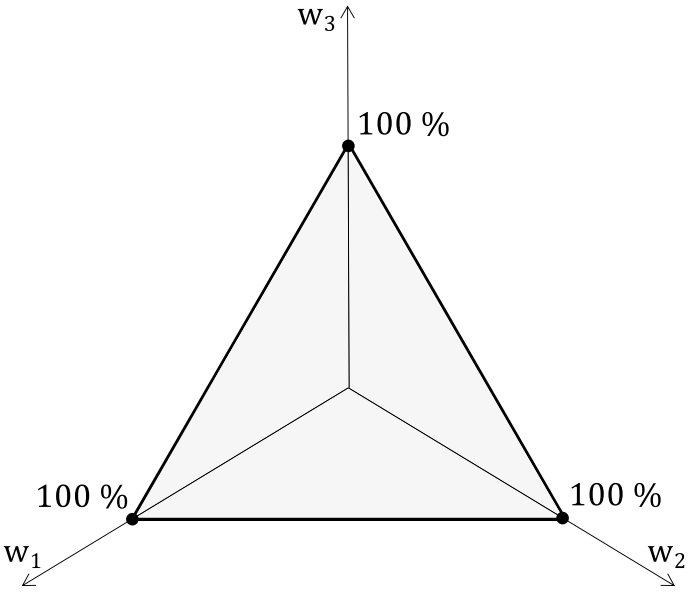


Fig. 2 Relative weight distributions among three voters in the unit simplex

5. For each weight distribution $\bar{\mathbf{w}}$, the six points in the simplex (cf. Fig. 2) that structurally correspond to $\bar{\mathbf{w}}$—that is, $(\bar{w}_1, \bar{w}_2, \bar{w}_3)$, $(\bar{w}_2, \bar{w}_1, \bar{w}_3)$, $(\bar{w}_1, \bar{w}_3, \bar{w}_2)$, etc.—are colored with the RGB intensities given by (R, G, B). For instance, $\mathcal{PBI}$ figures of $(1, 0, 0)$ translate into $(R, G, B) = (0, 0, 1)$ and dark blue color.

It is noteworthy that the distribution of voting power in two s-scoring committees $(N, A, r^s|\mathbf{w})$ and $(N, A, r^s|\mathbf{w}')$ can coincide even though the committees are non-equivalent: players are exactly as influential in either but some preference profiles yield different decisions so that $r^s|\mathbf{w} \not\equiv r^s|\mathbf{w}'$. Some of the illustrations in Fig. 3 therefore involve fewer different colors than there are distinct equivalence classes for the considered value of s. Moreover, equivalence classes that are represented by a single point in the simplex like the symmetric distribution of relative voting weights $\bar{\mathbf{w}} = (1/3, 1/3, 1/3)$, or a line—e.g., $\bar{\mathbf{w}} = (x, 1 - x, 0)$ for $0 < x < 1/2$—may not be visible without magnification. We have manually enlarged them only for $s = 0$ and $s = 1$. Bearing these caveats in mind, the colored simplices below provide accurate maps of all equivalence classes of scoring committees that exist for a given value of s.

For instance, the large blue triangles inside the panel for $s = 0$, i.e., weighted plurality committees, correspond to $\tilde{\mathbf{w}} = (1, 0, 0)$, i.e., the dictatorship of one player. The green midpoints of the simplex's boundary lines represent the equivalence class with $\tilde{\mathbf{w}} = (1, 1, 0)$: two players decide symmetrically, and the third never makes a difference. The simplex's light yellow midpoint reflects $\tilde{\mathbf{w}} = (1, 1, 1)$, i.e., three absolutely symmetric players. The remaining three plurality equivalence classes with reference weights of $\tilde{\mathbf{w}} = (2, 1, 1)$, $\tilde{\mathbf{w}} = (2, 2, 1)$, and $\tilde{\mathbf{w}} = (3, 2, 2)$ correspond, respectively,

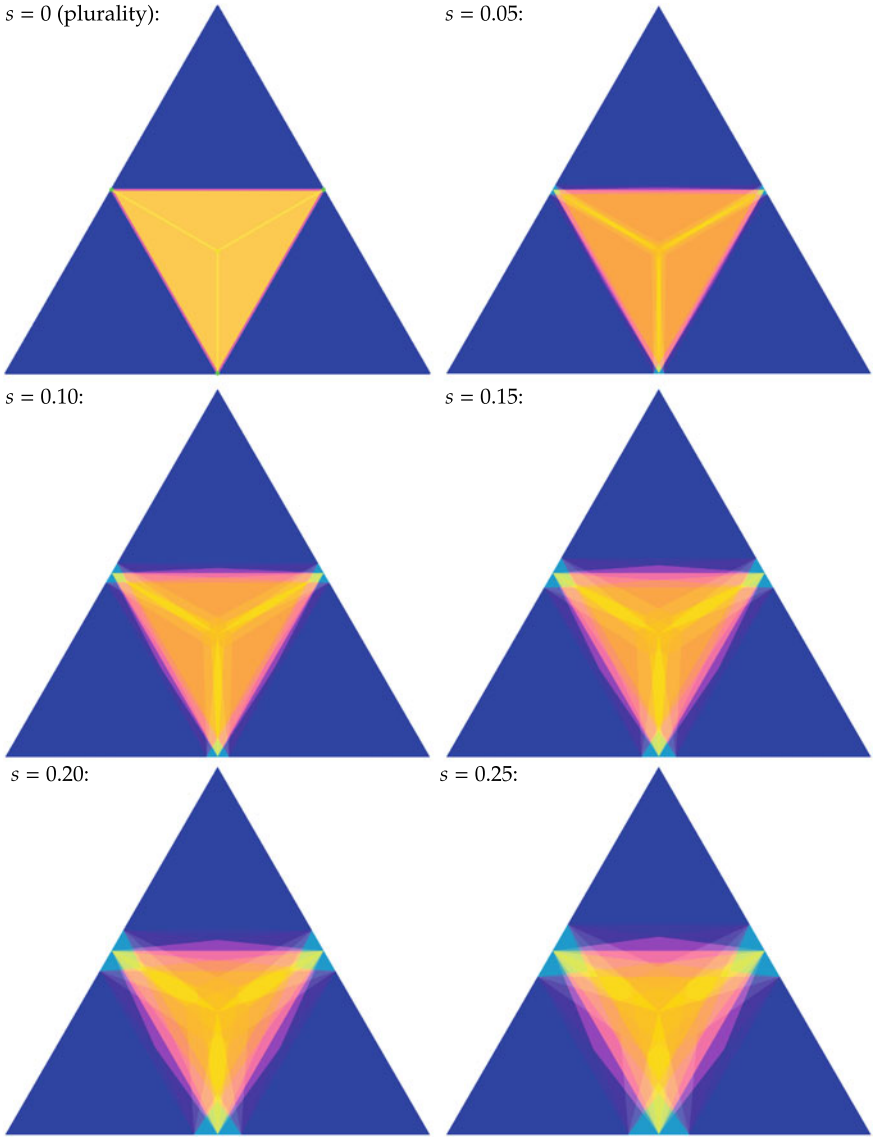


Fig. 3 Weighted s-scoring committees in Penrose-Banzhaf coloring

to the purple lines between the boundary midpoints, the dark yellow lines from the simplex's center to the three boundary midpoints, and the residual orange triangles. Lists of reference weight distributions for other values of s are provided by Mayer and Napel (2021).

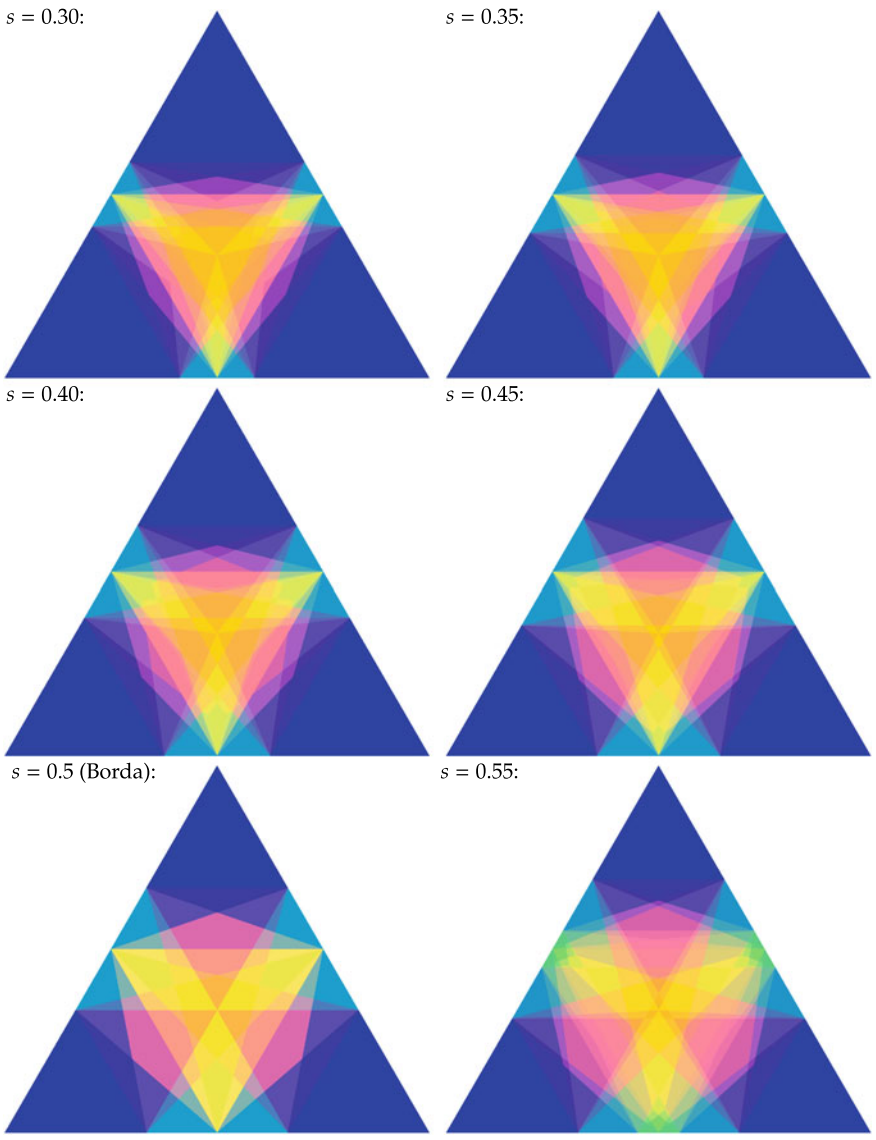


Fig. 3 (continued)

5 Concluding Remarks

The illustrations in Sect. 4 exhibit the hidden beauty of weighted voting in committees. However, the artistically appealing (at least to us) geometry and changing colors have substantive implications. They reveal structural properties of collective decision making in politics and economics.

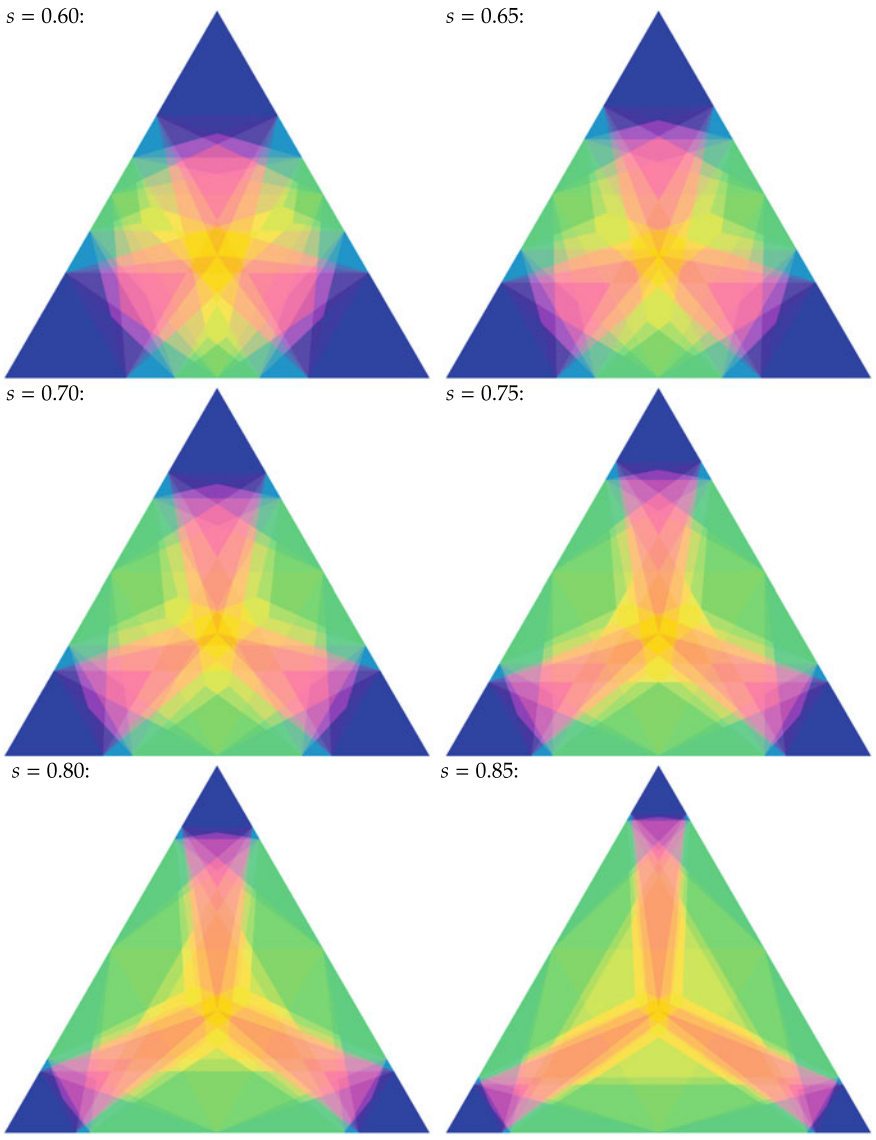


Fig. 3 (continued)

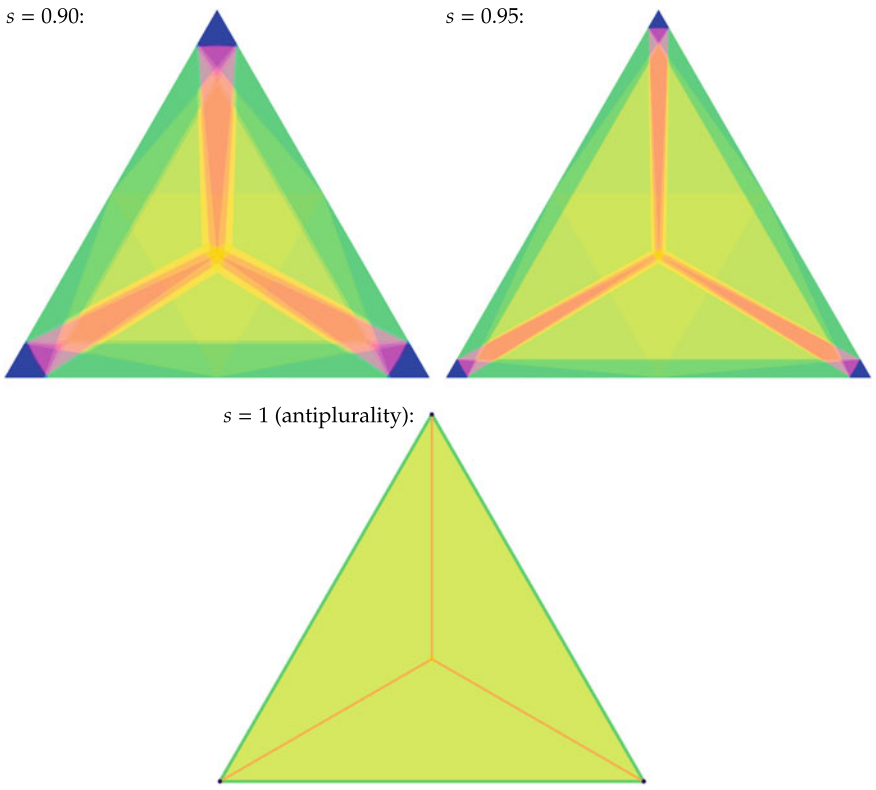


Fig. 3 (continued)

Take, for instance, the large blue triangles in the panel for $s = 0$. As pointed out already, they correspond to the dictatorship of one player. Namely, if some player's voting weight slightly exceeds 50%—because a shareholder has acquired a small majority stake in a corporation, committee seats are awarded in proportion to population shares in an ethnically polarized society with one majority and several large minorities, etc.—then all plurality decisions correspond to the top preference of that player. This is not the case for many other scoring rules: consider different levels of s and watch how the blue triangle shrinks from panel to panel. The shrinkage documents how basing decisions on more than just the top preference over all candidates makes the collective choice more 'inclusive'. For instance, adopting Borda's rule instead of taking plurality decisions turns a previous dictator with $\bar{w}_i$ slightly above 50% into just a very dominant player. Player i can swing the joint decision for many but no longer for all preference configurations under Borda's rule. Antiplurality awards dictatorial influence not even to a player who has a perfect monopoly of votes. The player's relative weight of 100% makes it impossible for the respec-

tive worst-ranked candidate to win but with lexicographic tie-breaking relatively few perturbations of the dominant player's preferences alter the winner.

The changing variety of colors in the panels visualizes the findings reported in Fig. 1: equivalence maps for scoring rules with $0 < s < 1/2$ or $1/2 < s < 1$ involve many more color shades than those for $s = 0$, $s = 1$, and also $s = 1/2$. We can moreover locate the ranges of weights where most of the color changes are concentrated, i.e., where sensitivity to small weight changes is the greatest. In these weight regions, the incentives to, say, increase one's corporate shareholdings or to try to attract a party switcher are much greater than in monochrome areas.

The multiplicity of colors in a panel also indicates the scope for achieving a particular distribution of influence as an institutional designer who can determine the distribution of weights (the so-called 'inverse problem' of voting power; cf. Kurz 2012). Think of a federation of three differently sized states: it may be desirable to make states' voting power a specific function of population sizes—e.g., to achieve direct proportionality or proportionality to the square root of population sizes. Though perfect symmetry (light yellow) or dictatorship by one state (dark blue) are always feasible, the chances of finding voting weights that achieve the targeted distribution of influence are arguably smaller for, say, $s = 1$ with only five equivalence classes than for $s = 1/2$ with 51. There is also a tendency for the distribution of relative voting power to match the underlying distribution of relative voting weights better, the more equivalence classes or colors in our illustrations. This relates to the so-called 'transparency' of a voting rule (cf. Kurz et al. 2021, Sect. 7.4).

We readily admit that illustrations of voting power in three-player committees that decide between three options have neither the complexity nor the aesthetic qualities of Julia sets or Mandelbrot sets, which have crossed the boundaries between art and science much earlier (see, e.g., Peitgen and Richter 1986). But there is definitely more art and beauty in weighted voting and the resulting voting power than typically meets the eye.

References

Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19*(2), 317–343.

Holler, M. J., & Nurmi, H. (Eds.). (2013). *Power, Voting, and Voting Power: 30 Years After*. Springer.

Holler, M. J., & Packel, E. W. (1983). Power, luck and the right index. *Zeitschrift für Nationalökonomie (Journal of Economics), 43*(1), 21–29.

Kovacic, M., & Zoli, C. (2021). Ethnic distribution, effective power and conflict. *Social Choice and Welfare, 57*, 257–299.

Kurz, S. (2012). On the inverse power index problem. *Optimization, 61*(8), 989–1011.

Kurz, S., Mayer, A., & Napel, S. (2020). Weighted committee games. *European Journal of Operational Research, 282*(3), 972–979.

Kurz, S., Mayer, A., & Napel, S. (2021). Influence in weighted committees. *European Economic Review, 132*, 103634.

Maskin, E., & Sen, A. (2016). How majority rule might have stopped Donald Trump. *New York Times*, April 28, 2016.

Mayer, A., & Napel, S. (2020). Weighted voting on the IMF Managing Director. *Economics of Governance, 21*(3), 237–244.
Mayer, A., & Napel, S. (2021). Weighted scoring committees. *Games, 12*(4), 94.
Napel, S. (2019). Voting power. In R. Congleton, B. Grofman, & S. Voigt (Eds.) *Oxford Handbook of Public Choice* (Vol. 1, Chap. 6, pp. 103–126). Oxford University Press.
Napel, S., & Welter, D. (2021). Simple voting games and cartel damage proportioning. *Games, 12*(4), 74.
Peitgen, H.-O., & Richter, P. H. (1986). *The Beauty of Fractals*. Springer.
Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society, 109*(1), 53–57.
Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*(3), 787–792.
Taylor, A. D., & Zwicker, W. S. (1999). *Simple Games*. Princeton University Press.
Von Neumann, J., & Morgenstern, O. (1953). *Theory of Games and Economic Behavior* (3rd ed.). Princeton University Press.

An Application of Power Indices for the Family of Weighted Majority Games in Partition Function Form

J. M. Alonso-Meijide, L. M. Armijos-Toro, B. Casas-Méndez, and M. A. Mosquera

1 Introduction

Game theory is a mathematical discipline that is dedicated to the study of decision problems in which various players interact. In a cooperative game, the players have mechanisms that allow them to take binding agreements. One of the main lines of research in cooperative game theory is the study of simple games and power indices. In a simple game, the worth of a coalition is 1 or 0, and then, there are two types of coalitions: winning and losing. Simple games are supposed to be monotonic, in the sense that the enlargement of a winning coalition cannot cause it to become a losing one. Frequently, simple games model voting situations. A power index provides a measure of a voter's ability to change the outcome of a voting. Many different power indices have been defined and studied. The most well-known are the Shapley–Shubik power index (1954) and the Banzhaf power index (1964). Both power indices assign to a player a measure that is based on the contributions the player makes by joining

J. M. Alonso-Meijide (✉)
MODESTYA Research Group, Department of Statistics, Mathematical Analysis and Optimisation, Faculty of Sciences, University of Santiago de Compostela, Campus de Lugo, 27002 Lugo, Spain
e-mail: josemaria.alonso@usc.es

L. M. Armijos-Toro
Departamento de Ciencias Exactas, Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
e-mail: livinomanuel.armijos@rai.usc.es

L. M. Armijos-Toro · B. Casas-Méndez
MODESTYA Research Group, Department of Statistics, Mathematical Analysis and Optimisation, Faculty of Mathematics, University of Santiago de Compostela, Campus Vida, 15782 Santiago de Compostela, Coruna, Spain
e-mail: balbina.casas.mendez@usc.es

M. A. Mosquera
Universidade de Vigo, Departamento de Estatística e Investigación Operativa, 32004 Ourense, Spain
e-mail: mamrguez@uvigo.es

J. M. Alonso-Meijide · L. M. Armijos-Toro · B. Casas-Méndez · M. A. Mosquera
CITMAga, 15782 Santiago de Compostela, Coruna, Spain

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_8

other coalitions. A winning coalition is minimal if it becomes a losing one if any of its members cease to be part of it, that is, all the players of a minimal winning coalition are decisive. Deegan and Packel (1978) and Holler (1982) proposed two power indices that assign a measure of power to a player taking into account only the minimal winning coalitions to which he belongs: the Deegan–Packel power index and the Public Good power index, respectively. The Deegan–Packel power index is based on the assumption that all minimal winning coalitions are equally likely, and that within a coalition all its players are equally important. The Public Good power index does not take into account the number of players who are part of the minimal winning coalitions, but the cardinality of the set of minimal winning coalitions to which a player belongs. Other power indices that follow similar arguments are those proposed in Johnston (1978) and Álvarez-Mozos et al. (2015). The Johnston power index takes into account only those winning coalitions in which there are some decisive players. The power index proposed by Álvarez-Mozos et al. (2015) only takes into account winning coalitions that do not contain any null player.

A particular class of simple games is the family of weighted majority games. A simple game is a weighted majority if there is a set of weights for the players and a quota, such that a coalition is winning if and only if the sum of the weights of the players of the coalition (that is, the weight of the coalition) is not less than the quota. Colomer and Martínez (1995) proposed a power index that is specific for the family of weighted majority games. The Colomer–Martínez power index, like the Deegan–Packel or the Public Good power indices, is also based on minimal winning coalitions. This power index has in common with the Deegan–Packel power index that all the minimal winning coalitions are considered equally likely, but it does not consider that all the players are equally important. The relevance of the players of a minimal winning coalition is directly related with their weights. Barua et al. (2005) defined a different power index for the family of weighted majority games. Although this power index is based on the Banzhaf power index, when absolute majority is required, the power index defined in Barua et al. Barua et al. (2005) assigns power in a proportional way to the weight of each player, analogous to the way the Colomer–Martínez power index acts. In Armijos-Toro et al. (2021), a new power index for the family of weighted majority games was defined and axiomatically characterized. This new power index assigns to a player an amount that is defined considering ideas of both the Public Good and Colomer–Martínez power indices, because it takes into account: (i) the cardinality of the minimal winning coalitions to which he belongs and (ii) his weight in the majority game.

In addition to the simple games discussed above, two other relevant types of cooperative games are as follows: games in characteristic function form and games in partition function form. In a game in characteristic function form, a coalition gets an amount regardless of how the rest of the players organize. The characteristic function assigns a value to each subset of players. Note that simple games are a subclass of the games in characteristic function form. However, in the more general model of games in partition function form (Thrall and Lucas 1963), what a coalition gets depends on the arrangement of the rest of the players. The partition function assigns a value to each embedded coalition, that is, to each pair formed by a coalition (called

active coalition) and a partition of the players outside the active coalition. Returning to the voting situations modeled by the simple games, the model of games in partition function form has proven to be useful to represent situations with plurality rules. In these cases, if no candidate is supported by a qualified majority, that which gets more votes gets the government. Thus, in a weighted majority game in partition function form, an embedded coalition is considered winning if the weight of the active coalition is greater than or equal to the weight of any of the coalitions that constitute the partition of the rest of the players. If two or more coalitions have the same greatest weight, that is, in case there are draws, to determine the active coalition for the winning embedded coalition a breaking rule (number of votes, for example) has to be used. In games in characteristic function form, a solution concept is intended to distribute the value of the total coalition among all players and two of the most prominent solution concepts are the so-called Shapley (1953) and Banzhaf values, respectively. Myerson (1977) studied a generalization of the Shapley value for games in partition function form. Dutta et al. (2010) studied a family of values that generalizes the Shapley value. de Clippel and Serrano (2008) proposed the externality-free Shapley value. Bolger (1983) and Bolger (1990) studied generalizations of the Banzhaf value for games in partition function form. In Álvarez-Mozos et al. (2017), the restriction of the externality-free Shapley value (the externality-free Shapley–Shubik power index) was axiomatically characterized. In Álvarez-Mozos and Tejada (2015), an alternative generalization of the Banzhaf value (the so-called ordinal Banzhaf power index) was proposed for the family of simple games in partition function form.

In Alonso-Meijide et al. (2017), a new class of simple games in partition function form was defined. Moreover, generalizations of the Deegan–Packel and Public Good power indices for these simple games in partition function form are proposed and characterized. To define these power indices, a notion of inclusion among embedded coalitions and a related property of monotonicity are introduced. Roughly speaking, an embedded coalition is larger when the active coalition does not decrease and the partition of the rest of the players becomes finer. Carreras and Magaña (2008) used a similar idea even though in a different setting.

There is also a large body of literature in which the methodology described above is applied to real problems, particularly in the field of politics, to measure the power of the different political parties represented in parliaments or other organizations. In Alonso-Meijide et al. (2017), the new power indices for simple games in partition function form there introduced were used to study the distribution of power in the Parliament of Andalusia, an autonomous community in Spain, that emerged after the elections of March 22, 2015. In Arévalo-Iglesias and Álvarez-Mozos (2020), some other well-known power indices extended to simple games in partition function form were applied to the Parliament of another autonomous community in Spain, the Basque Country. Armijos-Toro et al. (2021) presented an application of their new power index for the family of weighted majority games to the results of the elections held in 2021 for the formation of a new National Assembly of Ecuador. In this last case, the so-called legislative benches, also known as blocks, are the official political groupings within the assembly, with the right to have authority within the different legislative commissions. If initially, after the elections, there is a configuration of

these blocks, it may happen that several additional parliamentary groups are formed during the course of the legislative period, due to splits within the legislative benches, which raises the interest in analyzing the evolution of power within this Assembly over the course of a single legislative period.

In this paper, the Colomer–Martínez power index and the power index proposed in Armijos-Toro et al. (2021), are generalized for the model of weighted majority games in partition function form. They are also applied to the analysis of power within the National Assembly of Ecuador. To do this, we analyze minimal winning embedded coalitions in partition function form (plurality rule). Under plurality voting, the winning option is the one that received the highest number of votes, although this group has not necessarily the majority. Among the countries that use the plurality rule for elections are Canada and Great Britain. We use the National Assembly of Ecuador as a real-world numerical example to illustrate the computation of the proposed indices and compare their results.

The rest of the paper is organized as follows. In Sect. 2, we introduce some preliminaries and the basic notation for weighted majority games in partition function form. In Sect. 3, we generalize to this context of games in partition function form the Colomer–Martínez power index and the power index proposed in Armijos-Toro et al. (2021). Finally, in Sect. 4, a real example from the political field, the National Assembly of Ecuador, is used to illustrate the new power indices. To complete our study, we compute several power indices from the literature and compare the results.

2 Simple Games in Characteristic and Partition Function Form

In this section, we introduce the notation and preliminary concepts necessary for our study focusing on weighted majority games in partition function form.

2.1 Simple Games in Characteristic Function Form

A game in characteristic function form is a pair (N, v), where N is the finite set of players and $v : 2^N \longrightarrow \mathbb{R}$ is the characteristic function of the game satisfying $v(\emptyset) = 0$. Non-empty subsets of N are called coalitions. The number $v(S)$ is to be understood as the worth of coalition S regardless of how the player of $N \setminus S$ organizes. A game (N, v) is monotone when $v(S) \leq v(T)$ for all $S \subseteq T \subseteq N$. A simple game is a monotone game such that $v(N) = 1$ and, for every $S \subseteq N, v(S) = 0$ or $v(S) = 1$. We denote by $\mathcal{SI}$ the class of simple games. Let $(N, v) \in \mathcal{SI}$, then a coalition $S \subseteq N$ is a winning (resp. losing) coalition when $v(S) = 1$ ($v(S) = 0$). $W(v)$ denotes the set of winning coalitions for the game (N, v). A simple game (N, v) is called decisive whenever $S \in W(v)$ if, and only if, $N \setminus S \notin W(v)$ for each

$S \subseteq N$. A winning coalition $S \subseteq N$ is a minimal winning coalition if there exists no $T \subset S$ such that T is a winning coalition. We denote by $M(v)$ the set of minimal winning coalitions of the game (N, v) and let $M_i(v) = \{S \in M(v) : i \in S\}$ for each $i \in N$.

Let N be a set of players and $S \subseteq N$, $S \neq \emptyset$. The game $(N, u_S) \in \mathcal{SI}$ (called unanimity game of S) is the (non-decisive) simple game such that $M(u_S) = \{S\}$, that is, for every $T \subseteq N$,

$$u_S(T) = \begin{cases} 1 & \text{if } S \subseteq T \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

It is well known that $\left\{(N, u_S) : S \in 2^N \backslash \emptyset\right\}$ constitutes a basis of the vector space of games in characteristic function form with a set of players N.

Let $(N, v), (N, v') \in \mathcal{SI}$. Their union is the game $(N, v \vee v') \in \mathcal{SI}$ such that, for all $S \subseteq N$, $S \in W(v \vee v')$ if $S \in W(v)$ or $S \in W(v')$; or in an equivalent way, $(v \vee v')(S) = \max\{v(S), v'(S)\}$, for every $S \subseteq N$. Two games $(N, v), (N, v') \in \mathcal{SI}$ are mergeable if for all pairs of coalitions $S, T \subseteq N$ such that $S \in M(v)$ and $T \in M(v')$, it holds that $S \nsubseteq T$ and $T \nsubseteq S$. If $(N, v), (N, v') \in \mathcal{SI}$ are mergeable, then $M(v) \cap M(v') = \emptyset$ and $M(v \vee v') = M(v) \cup M(v')$.

A power index for the family of simple games in characteristic function form is a mapping, f, that assigns to every simple game $(N, v) \in \mathcal{SI}$ a vector $f(N, v) \in \mathbb{R}^N$, where $f_i(N, v)$ describes the power of agent i in the game (N, v). Throughout this paper, we only consider efficient power indices, that is, $\sum_{i \in N} f_i(N, v) = 1$.

The definitions of two power indices, the Deegan–Packel power index (Deegan and Packel 1978) and the Public Good power index (Holler 1982), have in common that they only take into account the minimal winning coalitions of the game are presented below.

Definition 2.1 The Deegan–Packel power index (DP) is the power index defined for every $(N, v) \in \mathcal{SI}$ and $i \in N$ by

$$DP_i(N, v) = \frac{1}{|M(v)|} \sum_{S \in M_i(v)} \frac{1}{|S|}.$$

Definition 2.2 The Public Good power index (PG) is the power index defined for every $(N, v) \in \mathcal{SI}$ and $i \in N$ by

$$PG_i(N, v) = \frac{|M_i(v)|}{\sum_{j \in N} |M_j(v)|}.$$

Although both power indices use only minimal winning coalitions, for the computation of the DP power index, the number of players that are part of each minimal winning coalition is taken into account, while, for the PG power index only, the

number of minimal winning coalitions to which a player belongs is used, regardless of their size.

A null player in a game $(N, v) \in \mathcal{SI}$ is a player $i \in N$ such that $M_i(v) = \emptyset$. Two players $i, j \in N$ are symmetric in a game $(N, v) \in \mathcal{SI}$ if, for all coalition $S \subseteq N \backslash \{i, j\}$ such that $S \notin W(v)$, $S \cup \{i\} \in W(v)$ if and only if $S \cup \{j\} \in W(v)$. A power index satisfies the property of null player if $f_i(N, v) = 0$ when i is a null player in (N, v). A power index satisfies the property of symmetry if $f_i(N, v) = f_j(N, v)$ when $i, j \in N$ are symmetric in (N, v). Both power indices, DP and PG, satisfy the null player and symmetry properties. Moreover, DP and PG power indices of a union game $(N, v \vee v')$ of two mergeable games (N, v) and (N, v') can be computed as a weighted sum of the corresponding power indices of the games (N, v) and (N, v'), with different weights in each case (see Deegan and Packel 1978; Holler and Packel 1983). For every pair of mergeable simple games $(N, v), (N, v') \in \mathcal{SI}$, it holds that

$$DP(N, v \vee v') = \frac{|M(v)|DP(N, v) + |M(v')|DP(N, v')}{|M(v \vee v')|} \text{ and}$$

$$PG(N, v \vee v') = \frac{\sum_{j \in N} |M_j(v)|PG(N, v) + \sum_{j \in N} |M_j(v')|PG(N, v')}{\sum_{j \in N} |M_j(v \vee v')|}.$$

In this paper, we study the family of weighted majority games, a particular class of simple games.

Definition 2.3 A simple game $(N, v) \in \mathcal{SI}$ is a weighted majority game if there exists a non-negative vector of weights, $\boldsymbol{w} = (w_1, w_2, ..., w_n)$, where w_i is the weight of player $i \in N$, and a quota $q > 0$ such that a coalition $S \in W(v)$ if and only if $w(S) = \sum_{i \in S} w_i \geq q$.

We denote by $\mathcal{SIW}$ the class of weighted majority games. We identify a weighted majority game by the tuple $[q; w_1, w_2, ..., w_n]$, or $[q; \boldsymbol{w}]$ when no confusion is possible. A majority game is a weighted majority game such that $w_i = 1$, for all $i \in N$. If n is an odd number, then the majority game $[(n + 1)/2; 1, ..., 1]$ is decisive.

Colomer and Martínez (1995) defined a power index for the family of weighted majority games. This power index could be considered as a non-symmetric generalization of the Deegan–Packel power index.

Definition 2.4 The Colomer–Martínez power index (CM) is the power index defined for every $[q; \boldsymbol{w}] \in \mathcal{SIW}$ and $i \in N$ by

$$CM_i(q; \boldsymbol{w}) = \frac{1}{|M(q; \boldsymbol{w})|} \sum_{S \in M_i(q; \boldsymbol{w})} \frac{w_i}{\sum_{j \in S} w_j}.$$

In Armijos-Toro et al. (2021), the HCM power index, a new power index for weighted majority games, was proposed. The definition of the HCM power index

shares ideas with the Colomer–Martínez and Public Good power indices. Like the Public Good power index, the total number of minimal winning coalitions to which a player belongs is considered, and like the Colomer–Martínez power index, his weight is taken into account to define the power index.

Definition 2.5 The Holler–Colomer–Martínez power index is the power index defined for every $[q; \boldsymbol{w}] \in \mathcal{SIW}$ and $i \in N$ by

$$HCM_i(q; \boldsymbol{w}) = \frac{|M_i(q; \boldsymbol{w})| w_i}{\sum_{j \in N} |M_j(q; \boldsymbol{w})| w_j}.$$

The HCM power index computes the power of each player as the proportion, balanced by its weights in the game, of the number of minimal winning coalitions to which it belongs. An alternative expression for the HCM power index of a player $i \in N$ in a simple game $[q; \boldsymbol{w}] \in \mathcal{SIW}$ is

$$HCM_i(q; \boldsymbol{w}) = \frac{\sum_{S \in M_i(q;\boldsymbol{w})} w_i}{\sum_{j \in N} \sum_{S \in M_j(q;\boldsymbol{w})} w_j} = \sum_{S \in M_i(q;\boldsymbol{w})} \frac{w_i}{\sum_{T \in M(q;\boldsymbol{w})} \sum_{j \in T} w_j}.$$

When the Deegan–Packel and the Public Good power indices are restricted to the family of weighted majority games, the main difference between these indices and the Colomer–Martínez and the HCM power indices is that the last two ones are not symmetric. In a simple game with a unique minimal winning coalition S, the Deegan–Packel and the Public Good power indices assign the same power to all players in S, but the Colomer–Martínez and the HCM power indices assign to each player of S a power directly related with his weight. Formally, given $(N, v) \in \mathcal{SIW}$, determined by $[q; \boldsymbol{w}]$, and $S \in 2^N \backslash \emptyset$ with $M(v) = \{S\}$, then $DP_i(N, v) = PG_i(N, v) = 1/|S|$ for every $i \in S$; meanwhile, $CM_i(q; \boldsymbol{w}) = HCM_i(q; \boldsymbol{w}) = w_i / \sum_{j \in S} w_j$. Moreover, for every game $(N, v) \in \mathcal{SIW}$, being $M(v) = \{S\}$, every $i, j \in S$ are symmetric and then

$$DP_i(N, v) = PG_i(N, v) = DP_j(N, v) = PG_j(N, v),$$

$$CM_i(q; \boldsymbol{w}) w_j = CM_j(q; \boldsymbol{w}) w_i = HCM_i(q; \boldsymbol{w}) w_j = HCM_j(q; \boldsymbol{w}) w_i.$$

So, the following property is satisfied by the Colomer–Martínez and the HCM power indices. A power index on $\mathcal{SIW}$ satisfies the property of weighted symmetry if $w_j f_i(q; \boldsymbol{w}) = w_i f_j(q; \boldsymbol{w})$ when $i, j \in S$ and $M(q; \boldsymbol{w}) = \{S\}$. Moreover, Colomer–Martínez and the HCM power indices also satisfy some merging properties adapted to the class of weighted majority games (see Armijos-Toro et al. 2021).

2.2 Simple Games in Partition Function Form

A partition of a finite set N is a collection $\{P_1, P_2, \ldots, P_m\}$ such that $P_i \subseteq N$ for all $i \in N$, $\bigcup_{i \in N} P_i = N$, and $P_i \cap P_j = \emptyset$ for all pair $i, j \in N$ with $i \neq j$. We denote by $\mathcal{P}(N)$ the set of partitions of a finite set N. We assume that the empty set is an element of every partition, that is, $\emptyset \in \mathsf{P}$ for every $\mathsf{P} \in \mathcal{P}(N)$. A partition P is coarser than Q if every block $B \in \mathsf{Q}$ is included in some block $A \in \mathsf{P}$, i.e., $\mathsf{P} \supseteq \mathsf{Q}$. Then $(\mathcal{P}(N), \supseteq)$ is a lattice, called the partition lattice. With this ordering, the bottom element of the lattice is the finest partition $\{\{1\}, ..., \{n\}\}$, while the top element is the coarsest partition $\{N\}$. An embedded coalition of N is a pair $(S; \mathsf{P})$, where $\mathsf{P} \in \mathcal{P}(N)$ and $S \in \mathsf{P}$, is the active coalition in P. We denote by EC^N the set of embedded coalitions of N, i.e. $EC^N = \{(S; \mathsf{P}) : \mathsf{P} \in \mathcal{P}(N) \text{ and } S \in \mathsf{P}\}$. We say that a player $i \in N$ participates in an embedded coalition $(S; \mathsf{P}) \in EC^N$ if player i belongs to S. We simplify and write $S \cup i$ and $S \setminus i$ instead $S \cup \{i\}$ and $S \setminus \{i\}$, respectively. Given $\mathsf{P} \in \mathcal{P}(N)$ and $i \in N$, we denote by $\mathsf{P}(i)$ the element of P that contains i, i.e., $\mathsf{P}(i) \in \mathsf{P}$ and $i \in \mathsf{P}(i)$.

A game in partition function form is a pair (N, v), where N is the finite set of players and $v : EC^N \to \mathbb{R}$ is the partition function of the game satisfying $v(\emptyset; \mathsf{P}) = 0$ for every $\mathsf{P} \in \mathcal{P}(N)$. The number $v(S; \mathsf{P})$ is to be understood as the worth of coalition S when the players are organized according to P. In a game in partition function form, it is possible that $v(S; \mathsf{P}) \neq v(S; \mathsf{P}')$, for two pairs of partitions $\mathsf{P}, \mathsf{P}' \in \mathcal{P}(N)$ with $S \in \mathsf{P}$ and $S \in \mathsf{P}'$. The set of games in partition function form with a common set of players N is denoted by $\mathcal{G}^N$ and the set of games in partition function form with an arbitrary set of players is denoted by $\mathcal{G}$. It is easy to notice that $\mathcal{G}^N$ is a vector space over $\mathbb{R}$. Indeed, de Clippel and Serrano (2008) devised a basis of the vector space that generalizes the basis of games in characteristic function form that consists of unanimity games defined in Eq. (1). Given $(S; \mathsf{P}) \in EC^N$, with $S \neq \emptyset$, let $\left(N, e_{(S;\mathsf{P})}\right) \in \mathcal{G}$ be defined for every $(T; \mathsf{Q}) \in EC^N$ by

$$e_{(S;\mathsf{P})}(T; \mathsf{Q}) = \begin{cases} 1 & \text{if } S \subseteq T \text{ and } \forall T' \in \mathsf{Q}_{-T}, \exists S' \in \mathsf{P} \text{ such that } T' \subseteq S' \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\mathsf{Q}_{-T} \in \mathcal{P}(N \setminus T)$ denotes the partition $\mathsf{Q} \setminus \{T\}$. de Clippel and Serrano (2008) showed that $\left\{\left(N, e_{(S;\mathsf{P})}\right) : (S; \mathsf{P}) \in EC^N \text{ and } S \neq \emptyset\right\}$ constitutes a basis of $\mathcal{G}^N$.

In this paper, we are concerned with a subclass of $\mathcal{G}$ that generalizes simple games in characteristic function form as introduced by von Neumann and Morgenstern (1944). For doing so, Alonso-Meijide et al. (2017) developed a concept of monotonicity for games in partition function form. The intuition behind monotonic games is that the enlargement of a coalition cannot cause a decrease in its worth. Therefore, in order to generalize this idea, Alonso-Meijide et al. (2017) used a notion of inclusion for embedded coalitions that will be of key importance for its results and that it is implicitly formulated in Eq. (2).

Definition 2.6 Let N be a finite set and $(S;\mathsf{P}), (T;\mathsf{Q}) \in EC^N$. We define the inclusion among embedded coalitions as follows:

$$(S;\mathsf{P}) \sqsubseteq (T;\mathsf{Q}) \Longleftrightarrow S \subseteq T \text{ and } \forall T' \in \mathsf{Q}_{-T}, \exists S' \in \mathsf{P} \text{ such that } T' \subseteq S'.$$

For instance,

$$(\{1\};\{1\},\{2,3,4\}) \sqsubseteq (\{1,2\};\{1,2\},\{3,4\}) \text{ and}$$
$$(\{1,2\};\{1,2\},\{3,4\}) \sqsubseteq (\{1,2\};\{1,2\},\{3\},\{4\}).$$

Nevertheless $(\{1\};\{1\},\{2\},\{3\},\{4\})$ and $(\{1,2\};\{1,2\},\{3,4\})$ are not comparable. Note that whenever $S \neq \emptyset$, $(S;\mathsf{P}) \sqsubseteq (T;\mathsf{Q})$ if and only if $e_{(S;\mathsf{P})}(T;\mathsf{Q}) = 1$. According to the above definition, an embedded coalition $(S;\mathsf{P})$ is a subset of another embedded coalition $(T;\mathsf{Q})$ if $S \subseteq T$ and the partition of $N \setminus T$ defined as $\{R \setminus T : R \in \mathsf{P}\}$ is coarser than Q_{-T}.

Let us note that, intuitively, if $(S;\mathsf{P}) \sqsubseteq (T;\mathsf{Q})$, then the embedded coalition $(S;\mathsf{P})$ (i) has fewer members than $(T;\mathsf{Q})$, in the sense of inclusion, and (ii) faces more organized opposition, in the sense of non-members being partitioned more coarsely. Therefore, $(S;\mathsf{P})$ clearly faces higher hurdles than $(T;\mathsf{Q})$ to getting its will.

We introduce the class of simple games in partition function form for which we first extend the notion of monotonicity to the games in partition function form.

Definition 2.7 A game in partition function form $(N,v) \in \mathcal{G}$ is monotone when $v(S;\mathsf{P}) \leq v(T;\mathsf{Q})$ for all $(S;\mathsf{P}), (T;\mathsf{Q}) \in EC^N$ such that $(S;\mathsf{P}) \sqsubseteq (T;\mathsf{Q})$.

In a monotonic game in partition function form, players have incentives to enlarge coalitions but keeping the opposition less organized.

Definition 2.8 A game in partition function form $(N,v) \in \mathcal{G}$ is a simple game in partition function form if it satisfies:

(i) $v(S;\mathsf{P}) \in \{0,1\}$, for every $(S;\mathsf{P}) \in EC^N$.
(ii) $v(N;\{\emptyset,N\}) = 1$.
(iii) (N,v) is monotone.

An embedded coalition, $(S;\mathsf{P}) \in EC^N$, is winning if $v(S;\mathsf{P}) = 1$ and losing if $v(S;\mathsf{P}) = 0$. We denote by $\mathcal{SG}$ the set of simple games in partition function form.

The simple games in partition function form, as defined above, are the generalization of simple games in characteristic function form. First, each embedded coalition is either winning or losing. Second, the grand coalition $(N;\{\emptyset,N\})$ is always a winning coalition. Third, suppose that $(S;\mathsf{P}) \in EC^N$ is a winning embedded coalition, then $(T;\mathsf{Q})$ is a winning embedded coalition when $(S;\mathsf{P}) \sqsubseteq (T;\mathsf{Q})$, i.e., the game is monotone. The games that form the basis of de Clippel and Serrano (2008), see Eq. (2), are examples of simple games in partition function form.

In this paper, a particular class of embedded coalitions, the minimal winning embedded coalitions, play a very important role. Let $(N, v) \in \mathcal{SG}$. A winning embedded coalition, $(S; \mathsf{P}) \in EC^N$ $(v(S; \mathsf{P}) = 1)$ is a minimal winning embedded coalition if every proper subset of it is a losing embedded coalition, i.e., if $(T; \mathsf{Q}) \sqsubset (S; \mathsf{P})$ implies that $v(T; \mathsf{Q}) = 0$.[1] The set of minimal winning embedded coalitions of a simple game (N, v) in partition function form is denoted by $\mathcal{M}(v)$ and the subset of minimal winning embedded coalitions such that a given player $i \in N$ participates is denoted by $\mathcal{M}_i(v)$, i.e., $\mathcal{M}_i(v) = \{(S; \mathsf{P}) \in \mathcal{M}(v) : i \in S\}$. Taking into account the inclusion relation among embedded coalitions, a minimal winning embedded coalition $(S; \mathsf{P}) \in EC^N$ is a winning embedded coalition such that the active coalition S be of a minimum size and inactive coalitions $\mathsf{P} \setminus S$ to be of maximum size.

A player $i \in N$ is a null player in $(N, v) \in \mathcal{SG}$ if he does not participate in any minimal winning embedded coalition, i.e., $\mathcal{M}_i(v) = \emptyset$. Two players $i, j \in N$ are symmetric in $(N, v) \in \mathcal{SG}$ if exchanging the two players does not change the type of a coalition, i.e., if for every $(S; \mathsf{P}) \in EC^N$ such that $S \subseteq N \setminus \{i, j\}$,

$$v\left(S \cup i; \mathsf{P}_{-S,\mathsf{P}(i)} \cup \{S \cup i, \mathsf{P}(i) \setminus i\}\right) = 1 \Leftrightarrow$$
$$v\left(S \cup j; \mathsf{P}_{-S,\mathsf{P}(j)} \cup \{S \cup j, \mathsf{P}(j) \setminus j\}\right) = 1,$$

where $\mathsf{P}_{-S,\mathsf{P}(k)} = (\mathsf{P}_{-S})_{-\mathsf{P}(k)}$, for every $k \in \{i, j\}$.

In the same way that simple games in characteristic function form, a simple game in partition function form is completely determined by the set of minimal winning embedded coalitions. In a sense, all the relevant information of a simple game in partition function form is condensed in the set of minimal winning embedded coalitions. This fact is formally presented in the next result proved in Alonso-Meijide et al. (2017).

Proposition 2.1 *Let $C \subseteq EC^N$ be such that there is no relation with respect to $\sqsubseteq$ between any pair $(S; \mathsf{P}), (T; \mathsf{Q}) \in C$. Then, there exists a unique simple game in partition function form, (N, v), such that $\mathcal{M}(v) = C$.*

3 Power Indices for Weighted Majority Games in Partition Function Form

A power index for the family of simple games in partition function form is a mapping, f, that assigns to every simple game in partition function form $(N, v) \in \mathcal{SG}$ a vector $\mathsf{f}(N, v) \in \mathbb{R}^N$, where $\mathsf{f}_i(N, v)$ describes the power of agent i in the game (N, v).

In this paper, we are interested in the family of weighted majority games in partition function form, a particular class of simple games in partition function form.

[1] A proper subset, $(T; \mathsf{Q}) \sqsubset (S; \mathsf{P})$, is a subset $(T; \mathsf{Q}) \sqsubseteq (S; \mathsf{P})$ satisfying $(T; \mathsf{Q}) \neq (S; \mathsf{P})$.

Definition 3.1 A simple game in partition function form $(N, v) \in \mathcal{SG}$ is a weighted majority game in partition function form if there exists a non-negative vector of weights, $\boldsymbol{w} = (w_1, w_2, ..., w_n)$, where w_i is the weight of player $i \in N$, such that for every $(S; \mathsf{P}) \in EC^N$

$$v(S; \mathsf{P}) = 1 \iff \sum_{i \in S} w_i \geq \sum_{i \in T} w_i \quad \forall T \in \mathsf{P}.$$

We denote by $\mathcal{SWG}$ the set of weighted majority games in partition function form.

In the following, we extend the four power indices reviewed in Sect. 2.1 to this class of games.

Definition 3.2 DP is the power index defined for every $(N, v) \in \mathcal{SG}$ and $i \in N$ by

$$\mathsf{DP}_i(N, v) = \frac{1}{|\mathcal{M}(v)|} \sum_{(S;\mathsf{P}) \in \mathcal{M}_i(v)} \frac{1}{|S|}.$$

Definition 3.3 PG is the power index defined for every $(N, v) \in \mathcal{SG}$ and $i \in N$ by

$$\mathsf{PG}_i(N, v) = \frac{|\mathcal{M}_i(v)|}{\sum_{j \in N} |\mathcal{M}_j(v)|}.$$

The Colomer–Martínez and Holler–Colomer–Martínez power indices depend on the weights of each player, for this reason, we identify a weighted majority game in partition function form (N, v) by its vector of weights $\boldsymbol{w}$.

Definition 3.4 CM is the power index defined for every $(N, v) \in \mathcal{SWG}$, determined by the vector of weights $\boldsymbol{w}$, and $i \in N$ by

$$\mathsf{CM}_i(\boldsymbol{w}) = \frac{1}{|\mathcal{M}(\boldsymbol{w})|} \sum_{(S;\mathsf{P}) \in \mathcal{M}_i(\boldsymbol{w})} \frac{w_i}{\sum_{j \in S} w_j}.$$

Definition 3.5 HCM is the power index defined for every $(N, v) \in \mathcal{SWG}$, determined by the vector of weights $\boldsymbol{w}$, and $i \in N$ by

$$\mathsf{HCM}_i(\boldsymbol{w}) = \frac{|\mathcal{M}_i(\boldsymbol{w})| w_i}{\sum_{j \in N} |\mathcal{M}_j(\boldsymbol{w})| w_j}.$$

The two first extensions were studied and characterized in Alonso-Meijide et al. (2017). The properties used in the characterizations are the natural generalizations of the properties used in the original versions of the Deegan–Packel and Public Good power indices (see Deegan and Packel 1978; Holler 1982) for simple games in characteristic function form. To the best of our knowledge, the last two extensions are presented for the first time in this article.

Table 1 National Assembly of Ecuador in May 2021

Parties	Votes	Assembly members
UNES	5060922	49
MUPP	2530803	27
ID	1808867	18
PSC	1615833	18
CREO	1509436	12
IND	2061845	13

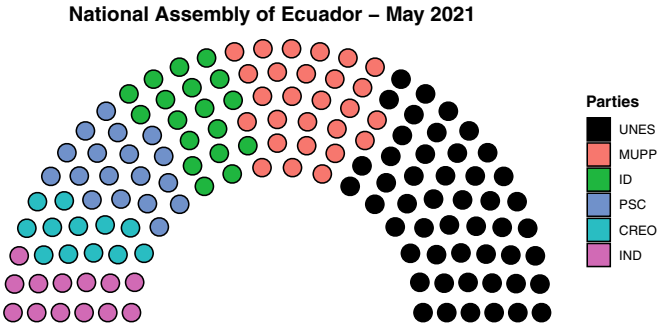


Fig. 1 National Assembly of Ecuador in May 2021

4 A Political Example: The National Assembly of Ecuador

In this section, we evaluate the results of the proposed power indices for weighted majority games in partition function form applied to the National Assembly of Ecuador. The analysis of this Assembly using weighted majority games in characteristic function form can be found in Armijos-Toro et al. (2021). The National Assembly of Ecuador consists of 137 assembly members. In February 2021, general elections were held in Ecuador.[2] The National Assembly was composed of: (49) UNES, (27) MUPP, (18) ID, (18) PSC, (12) CREO, and minorities (IND): (2) AVA, (2) MEU, (2) AH, (1) PSP, (1) AU, (1) MAP, (1) MUE, (1) MMI, (1) MAE, and (1) DEMSI. The parties CREO and PSC are pro-free market. Meanwhile, the parties ID and UNES are of progressive political ideology. Finally, the party MUPP is closer to socialism. The current president of Ecuador belongs to CREO party. The results of assembly members together with the votes obtained by each party are summarized in Table 1 and Fig. 1.

In June 2021, the party of the president of Ecuador (CREO) and some assembly members of the other parties, mainly from minorities, consolidated the new legislative bench (25) BAN. Also, some assembly members declared themselves independent:

[2] https://www.primicias.ec/noticias/politica/los-cambios-en-las-bancadas-de-la-asamblea/, last accessed 23/12/2021.

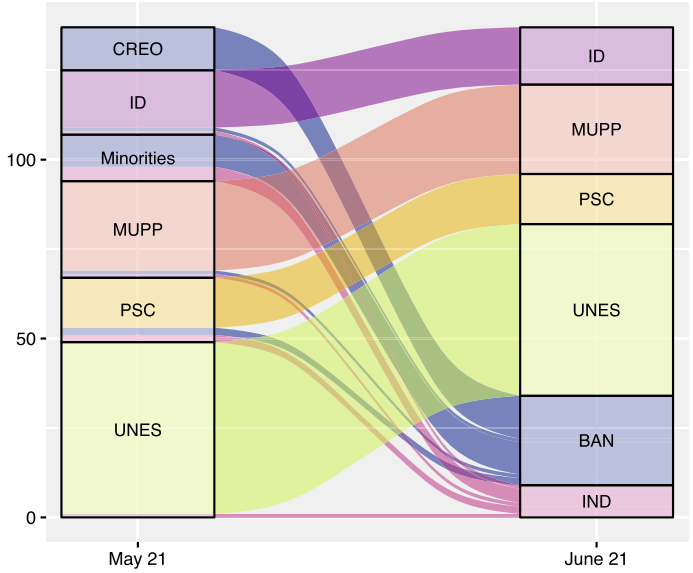


Fig. 2 Composition of the legislative benches—National Assembly of Ecuador in June 2021

(9) IND. Figure 2 shows the redistribution of assembly members in the legislative benches in June 2021.

In the 7 months of operation of the National Assembly of Ecuador, there have been changes in the structure of its legislative benches. These changes are shown in Table 2 and Fig. 3 and correspond to June 2021, July 2021,[3] 12 October 2021,[4] 26 October 2021,[5] and December 2021.[6]

Given that the National Assembly of Ecuador in May 2021 was structured in political parties and, as of June 2021, it was structured in legislative benches, we divided the analysis for May 2021 and a comparative analysis from June to December 2021.

We consider a decision procedure based on the plurality rule. If more of two options are possible, in the sense that players are organized in different groups, an option to win needs more votes than any other option. Therefore, the games we will consider will be weighted majority games in partition function form as defined in Definition 3.1. The weights correspond first to the original National Assembly

[3] https://www.primicias.ec/noticias/politica/bancadas-pierden-miembros-votos-asamblea/, last accessed 23/12/2021.

[4] https://www.primicias.ec/noticias/politica/union-unes-pachakutik-debilidad-legislativa-gobierno/, last accessed 23/12/2021.

[5] https://www.primicias.ec/noticias/politica/posible-destitucion-lasso-apoyo-asamblea/, last accessed 23/12/2021.

[6] https://www.primicias.ec/noticias/politica/ruptura-pachakutik-capitulo-bancadas-desgranads/, last accessed 23/12/2021.

Table 2 Changes in the legislative benches in the National Assembly of Ecuador

Benches	Jun 21	Jul 21	12 Oct 21	26 Oct 21	Dec 21
UNES	48	47	47	47	47
MUPP	25	24	25	25	25
BAN	25	25	25	26	28
ID	16	16	14	14	14
PSC	14	14	14	14	14
IND	9	11	12	11	9

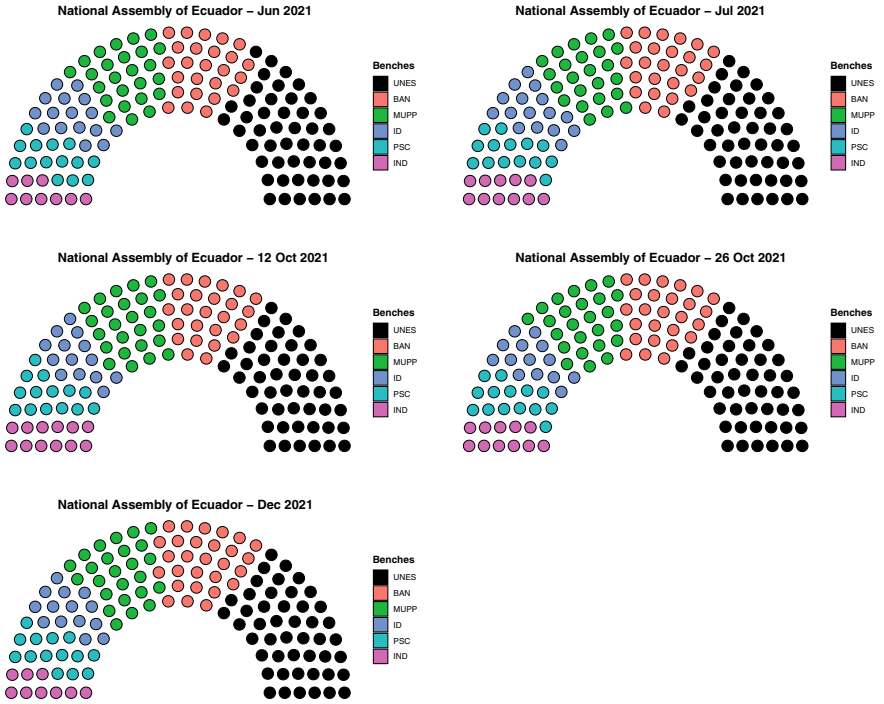


Fig. 3 National Assembly of Ecuador from June to December 2021

of Ecuador on May 2021 and subsequently those corresponding to the different legislative benches made, from June 2021 to December 2021. It should be kept in mind when defining the winning embedded coalitions that, in case of ties between coalitions with the highest overall weight, a tie-breaking rule should be used to determine the winner of an Assembly. One proposal is to count the number of votes in the elections and this criterion is the one we are going to use for May 2021. This tie-breaking rule was already used in Alonso-Meijide et al. (2017) and in Arévalo-Iglesias and Álvarez-Mozos (2020). In the cases corresponding to June, October, and December 2021 this tie-breaking rule is not applicable since the legislative benches

do not coincide exactly with the parties that presented themselves. So, one possibility is to consider that in case of a tie of two coalitions, both are winners, as proposed in van den Brink et al. (2021). However, we will use the criterion that, in the event of a tie, none of the coalitions is considered the winner. This criterion is used in many voting systems: in the event that there is no single most-voted option, the voting has to be repeated. In the case of the National Assembly of Ecuador, there are few ties and either of the above two options for breaking ties leads to very similar results.

4.1 Original Formation of the National Assembly of Ecuador: May 2021

First, we analyze the initial situation of the National Assembly of Ecuador in May 2021. In the case of ties, the vote received by each of the parties in the February 2021 general elections is used. In Table 3, we observe the 34 minimal winning embedded coalitions for the National Assembly of Ecuador in May 2021. Table 4 shows the two ties found in determining the winning embedded coalitions. In this table, the last four columns show the number of seats of the different coalitions that make up the partition. As mentioned above, the active coalition for these partitions will be the one that, having the highest number of assembly members, has also obtained the highest number of votes in the February 2021 general elections.

In the initial situation of the National Assembly of Ecuador in May 2021, we observe that UNES is present in 22 of the 34 active coalitions. Therefore, UNES is the party with the most power in the National Assembly of Ecuador in May 2021. UNES is also the only party capable of having active coalitions of cardinality one (9 of 34). Likewise, UNES is present in all active coalitions of cardinality two (12). The party of the president of Ecuador, CREO, is present in 10 of the 34 active coalitions.

Table 5 presents the results of the calculation of different power indices for the game in partition function form (plurality rule) in the National Assembly of Ecuador in May 2021. UNES is the party with the most power in the National Assembly of Ecuador in May 2021. Note that, although MUPP has more assembly members than ID and PSC, according to the Deegan–Packel power index, ID and PSC have more power than MUPP. This is because ID and PSC are present in 3 of the 12 active coalitions of cardinality two, while MUPP is only present in 1 of these 12 active coalitions. MUPP, ID, and PSC are each present in 11 of the 34 minimal winning embedded coalitions. Of the four proposed power indices, the Public Good power index gives the least power to UNES, but it still has the most power. This is because only the cardinality of the set of minimal winning embedded coalitions to which each player belongs is taken into account. Recall that UNES is present in 22 out of 34 minimal winning embedded coalitions, while the rest of the legislative benches are present in 10 or 11 of them.

Table 3 Minimal winning embedded coalitions—National Assembly of Ecuador in May 2021

Active coalition	Partition
{UNES}	{ID, PSC, CREO}, {MUPP, IND}, {UNES}
{UNES}	{ID, CREO, IND}, {MUPP, PSC}, {UNES}
{UNES}	{PSC, CREO, IND}, {MUPP, ID}, {UNES}
{UNES}	{MUPP, CREO}, {ID, IND}, {UNES}, {PSC}
{UNES}	{MUPP, CREO}, {ID, PSC}, {UNES}, {IND}
{UNES}	{MUPP, CREO}, {PSC, IND}, {UNES}, {ID}
{UNES}	{ID, IND}, {PSC, CREO}, {UNES}, {MUPP}
{UNES}	{ID, PSC}, {CREO, IND}, {UNES}, {MUPP}
{UNES}	{ID, CREO}, {PSC, IND}, {UNES}, {MUPP}
{UNES, MUPP}	{ID, PSC, CREO, IND}, {UNES, MUPP}
{UNES, ID}	{MUPP, PSC, IND}, {UNES, ID}, {CREO}
{UNES, ID}	{MUPP, CREO, IND}, {UNES, ID}, {PSC}
{UNES, ID}	{MUPP, PSC, CREO}, {UNES, ID}, {IND}
{UNES, PSC}	{MUPP, ID, IND}, {UNES, PSC}, {CREO}
{UNES, PSC}	{MUPP, ID, CREO }, {UNES, PSC}, {IND}
{UNES, PSC}	{MUPP, CREO, IND}, {UNES, PSC}, {ID}
{UNES, CREO}	{MUPP, ID, IND}, {UNES, CREO}, {PSC}
{UNES, CREO}	{ID, PSC, IND}, {UNES, CREO}, {MUPP}
{UNES, CREO}	{MUPP, PSC, IND}, {UNES, CREO}, {ID}
{UNES, IND}	{MUPP, PSC, CREO}, {UNES, IND}, {ID}
{UNES, IND}	{MUPP, ID, CREO}, {UNES, IND}, {PSC}
{UNES, CREO, IND}	{UNES, CREO, IND}, {MUPP, ID, PSC}
{MUPP, ID, PSC}	{MUPP, ID, PSC}, {UNES, CREO}, {IND}
{MUPP, ID, PSC}	{MUPP, ID, PSC}, {UNES, IND}, {CREO}
{MUPP, ID, PSC}	{MUPP, ID, PSC}, {CREO, IND}, {UNES}
{MUPP, ID, CREO}	{MUPP, ID, CREO}, {PSC, IND}, {UNES}
{MUPP, PSC, CREO}	{MUPP, PSC, CREO}, {ID, IND}, {UNES}
{MUPP, ID, IND}	{MUPP, ID, IND}, {PSC, CREO}, {UNES}
{MUPP, PSC, IND}	{MUPP, PSC, IND}, {ID, CREO}, {UNES}
{MUPP, CREO, IND}	{MUPP, CREO, IND}, {ID, PSC}, {UNES}
{ID, PSC, IND}	{ID, PSC, IND}, {MUPP, CREO}, {UNES}
{MUPP, PSC, CREO, IND}	{MUPP, PSC, CREO, IND}, {UNES, ID}
{MUPP, ID, CREO, IND}	{MUPP, ID, CREO, IND}, {UNES, PSC}
{ID, PSC, CREO, IND}	{ID, PSC, CREO, IND}, {UNES}, {MUPP}

Table 4 Partitions that result in ties—National Assembly of Ecuador in May 2021

Partition	Seats P_1	Seats P_2	Seats P_3	Seats P_4
{ID, PSC, IND}, {MUPP, CREO}, {UNES}	49	39	49	–
{ID, PSC, IND}, {UNES}, {MUPP}, {CREO}	49	49	27	12

Table 5 Power indices—National Assembly of Ecuador in May 2021

Power indices	UNES	MUPP	ID	PSC	CREO	IND
DP	0.4510	0.1078	0.1176	0.1176	0.1054	0.1005
PG	0.2933	0.1467	0.1467	0.1467	0.1333	0.1333
CM	0.5496	0.1415	0.0944	0.0944	0.0572	0.0630
HCM	0.5334	0.1470	0.0980	0.0980	0.0594	0.0643

4.2 Formation of the National Assembly of Ecuador: From Jun 2021 to Dec 2021

Table 6 shows the 37 minimal winning embedded coalitions for the National Assembly of Ecuador in June 2021. Let us note that there are five ties in determining the winning embedded coalitions (see Table 7). As mentioned above, there is no winning embedded coalition in any of these partitions. UNES is present in 28 of the 37 minimal embedded coalitions and it is the only legislative bench in active coalitions of cardinality one. Note that, although BAN and MUPP have more assembly members than PSC and ID, the latter appear in more minimal winning embedded coalitions (9 and 11, respectively) than the former (8 each).

We could observe changes in the minimal winning embedded coalitions for each game from June to December 2021 in the National Assembly of Ecuador. In July 2021, two of the partitions that resulted in ties in June 2021 become minimal winning embedded coalitions, namely {{MUPP, PSC, IND}, {BAN, ID}, {UNES}} with active coalition {MUPP,PSC, IND}, and {{BAN, PSC, IND}, {MUPP, ID}, {UNES}} with active coalition {BAN, PSC, IND}. The cardinality of the set of minimal winning embedded coalitions is then 39. Table 8 shows the unique tie found in determining the winning coalitions in this case.

On 12 October 2021, the minimal winning embedded coalitions are 40. The winning embedded coalition {{UNES, ID, PSC}, {MUPP, BAN, IND}} with active coalition {UNES, ID, PSC} and the winning embedded coalition {{BAN, MUPP, PSC}, {UNES, ID}, {IND}} with active coalition {BAN, MUPP, PSC} (this partition resulted in tie in July 2021, see Table 8) become minimal winning embedded coalitions. Moreover, with respect to July 2021, the minimal winning embedded coalitions associated with partitions {{MUPP, BAN, IND}, {UNES, PSC},{ID}} and {{MUPP, BAN, IND}, {UNES, ID}, {PSC}} change their active coalitions to {MUPP, BAN, IND} in both cases. Lastly, the embedded coalition {{MUPP, BAN,

Table 6 Minimal winning embedded coalitions—National Assembly of Ecuador in June 2021

Active coalition	Partition
{UNES}	{MUPP, IND}, {BAN, PSC}, {UNES}, {ID}
{UNES}	{MUPP, IND}, {BAN, ID}, {UNES}, {PSC}
{UNES}	{MUPP, IND}, {ID, PSC}, {UNES}, {BAN}
{UNES}	{MUPP, ID}, {BAN, IND}, {UNES}, {PSC}
{UNES}	{MUPP, ID}, {BAN, PSC}, {UNES}, {IND}
{UNES}	{MUPP, ID}, {PSC, IND}, {UNES}, {BAN}
{UNES}	{MUPP, PSC}, {BAN, IND}, {UNES}, {ID}
{UNES}	{MUPP, PSC}, {BAN, ID}, {UNES}, {IND}
{UNES}	{MUPP, PSC}, {ID, IND}, {UNES}, {BAN}
{UNES}	{BAN, IND}, {ID, PSC}, {UNES}, {MUPP}
{UNES}	{BAN, ID}, {PSC, IND}, {UNES}, {MUPP}
{UNES}	{BAN, PSC}, {ID, IND}, {UNES}, {MUPP}
{UNES}	{ID, PSC, IND}, {UNES}, {MUPP}, {BAN}
{UNES, BAN}	{MUPP, ID, PSC, IND}, {UNES, BAN}
{UNES, MUPP}	{BAN, ID, PSC, IND}, {UNES, MUPP}
{UNES, PSC}	{MUPP, BAN, IND}, {UNES, PSC}, {ID}
{UNES, ID}	{MUPP, BAN, IND}, {UNES, ID}, {PSC}
{UNES, PSC}	{MUPP, ID, IND}, {UNES, PSC}, {BAN}
{UNES, IND}	{UNES, IND}, {MUPP, BAN}, {ID, PSC}
{UNES, ID}	{UNES, ID}, {MUPP, BAN}, {PSC, IND}
{UNES, PSC}	{UNES, PSC}, {MUPP, BAN}, {ID, IND}
{UNES, ID}	{BAN, PSC, IND}, {UNES, ID}, {MUPP}
{UNES, ID}	{MUPP, PSC, IND}, {UNES, ID}, {BAN}
{UNES, PSC}	{BAN, ID, IND}, {UNES, PSC}, {MUPP}
{UNES, IND}	{MUPP, ID, PSC}, {UNES, IND}, {BAN}
{UNES, IND}	{BAN, ID, PSC}, {UNES, IND}, {MUPP}
{MUPP, BAN}	{ID, PSC, IND}, {MUPP, BAN}, {UNES}
{UNES, ID, IND}	{UNES, ID, IND}, {MUPP, BAN, PSC}
{UNES, PSC, IND}	{UNES, PSC, IND},{MUPP, BAN, ID}
{MUPP, BAN, PSC}	{MUPP, BAN, PSC}, {UNES, IND}, {ID}
{MUPP, BAN, ID}	{MUPP, BAN, ID}, {UNES, PSC}, {IND}
{MUPP, BAN, ID}	{MUPP, BAN, ID}, {UNES, IND}, {PSC}
{MUPP, ID, IND}	{MUPP, ID, IND}, {BAN, PSC}, {UNES}
{MUPP, ID, PSC}	{MUPP, ID, PSC}, {BAN, IND}, {UNES}
{BAN, ID, IND}	{BAN, ID, IND}, {MUPP, PSC}, {UNES}
{BAN, ID, PSC}	{BAN, ID, PSC}, {MUPP, IND}, {UNES}
{MUPP, BAN, PSC, IND}	{MUPP, BAN, PSC, IND}, {UNES, ID}

Table 7 Partitions that result in ties—National Assembly of Ecuador in June 2021

Partition	Seats P_1	Seats P_2	Seats P_3	Seats P_4
{BAN, MUPP, PSC}, {UNES, ID}, {IND}	64	64	9	–
{BAN, PSC, IND}, {MUPP, ID}, {UNES}	48	41	48	–
{MUPP, PSC, IND}, {BAN, ID}, {UNES}	48	41	48	–
{BAN, PSC, IND}, {UNES}, {MUPP}, {ID}	48	48	25	16
{MUPP, PSC, IND}, {UNES}, {BAN}, {ID}	48	48	25	16

Table 8 Partition that results in tie—National Assembly of Ecuador in July 2021

Partition	Seats P_1	Seats P_2	Seats P_3
{BAN, MUPP, PSC}, {UNES, ID}, {IND}	63	63	11

Table 9 Partition that results in tie—National Assembly of Ecuador in December 2021

Partition	Seats P_1	Seats P_2	Seats P_3
{BAN, ID, PSC}, {UNES, IND}, {MUPP}	56	56	25

PSC, IND}, {UNES, ID}} with active coalition {MUPP, BAN, PSC, IND} ceases to be minimal since it contains the winning embedded coalition {{MUPP, BAN, IND}, {PSC}, {UNES, ID}} with active coalition {MUPP, BAN, IND}.

The minimal winning embedded coalitions on 26 October 2021 are the same as on 12 October 2021. There are no ties in either case. Finally, we observe only one change in December 2021. The minimal winning embedded coalition {{BAN, ID, PSC}, {UNES, IND}, {MUPP}} with active coalition {UNES, IND} is out since its partition results in tie (see Table 9). So, we have 39 minimal winning embedded coalitions in December 2021.

Table 10 shows the distribution of power of the different legislative benches in the National Assembly of Ecuador from June to December 2021. In the period of analysis, we observe that UNES is the legislative bench with the most power in the Assembly. The Deegan–Packel and Public Good power indices do not change the power assigned to MUPP and BAN in spite of changes observed in the structure of the legislative benches. Nevertheless, the Colomer–Martínez and HCM power indices show variations according to these changes.

Next, we provide some general comments. First, UNES is allocated with about half or more of the total power at almost all considered periods, with the exception of the Public Good power index. This may be due to the fact that, in the formation of winning embedded coalitions, the participation of UNES is preferable since it is the legislative bench with the highest number of votes. Second, with respect to the Deegan–Packel and Public Good power indices when the simple majority rule is used, these power indices do not show variation for the legislative benches in the period of analysis (see Armijos-Toro et al. 2021). This is because these power indices are considered as

Table 10 Power indices—National Assembly of Ecuador

Benches	Power indices	Jun 21	Jul 21	12 Oct 21	26 Oct 21	Dec 21
UNES	DP	0.5450	0.5171	0.4875	0.4875	0.4872
	PG	0.3889	0.3590	0.3293	0.3293	0.3250
	CM	0.6560	0.6179	0.5808	0.5821	0.5782
	HCM	0.6346	0.5990	0.5618	0.5618	0.5547
MUPP	DP	0.0788	0.0833	0.1000	0.1000	0.1026
	PG	0.1111	0.1154	0.1341	0.1341	0.1375
	CM	0.0889	0.0943	0.1167	0.1164	0.1188
	HCM	0.0944	0.0983	0.1217	0.1217	0.1248
BAN	DP	0.0788	0.0833	0.1000	0.1000	0.1026
	PG	0.1111	0.1154	0.1341	0.1341	0.1375
	CM	0.0889	0.0973	0.1167	0.1201	0.1301
	HCM	0.0944	0.1024	0.1217	0.1266	0.1398
ID	DP	0.1171	0.1111	0.1042	0.1042	0.1068
	PG	0.1528	0.1410	0.1341	0.1341	0.1375
	CM	0.0791	0.0752	0.0645	0.0645	0.0660
	HCM	0.0831	0.0801	0.0682	0.0682	0.0699
PSC	DP	0.0968	0.1090	0.1042	0.1042	0.1068
	PG	0.1250	0.1410	0.1341	0.1341	0.1375
	CM	0.0546	0.0668	0.0645	0.0645	0.0660
	HCM	0.0595	0.0701	0.0682	0.0682	0.0699
IND	DP	0.0833	0.0962	0.1042	0.1042	0.0940
	PG	0.1111	0.1282	0.1341	0.1341	0.1250
	CM	0.0326	0.0485	0.0567	0.0525	0.0409
	HCM	0.0340	0.0501	0.0584	0.0536	0.0409

valid the minimal winning coalitions, which do not change over the analyzed period. Whereas, when using the plurality rule, the minimal winning embedded coalitions change. Thus, the Deegan–Packel and Public Good power indices present updates in the distribution of power in the National Assembly of Ecuador. Third, the Colomer–Martínez and HCM power indices, specifically defined for weighted majority games, better represent the alterations made in the Assembly structure over the period of analysis.

Finally, another notable result is that ID, using the Deegan–Packel and Public Good power indices, has greater power than BAN or MUPP. This is despite the fact that ID has fewer assembly members than BAN or MUPP. However, ID participates in a larger number of minimal winning embedded coalitions than BAN or MUPP. As noted, the Deegan–Packel and Public Good power indices distribute power equally among minimal winning embedded coalitions. However, the Colomer–Martínez and

HCM power indices consider in their calculation the number of assembly members of each legislative bench.

5 Conclusions

While the study of games in partition function form and simple games in partition function form is more abundant, it is less so for the case of weighted majority games in partition function form. In this paper, we extend to this context two power indices axiomatically characterized in Armijos-Toro et al. (2021) in the context of classical weighted majority games, one of them introduced in Colomer and Martínez (1995) and the other defined in Armijos-Toro et al.'s own work by combining the ideas of Colomer and Martínez (1995) and Holler (1982). Both extensions are illustrated by means of the case of the National Assembly of Ecuador in 2021 and the results obtained are compared with two power indices proposed and axiomatically characterized in Alonso-Meijide et al. (2017) in the context of simple games in partition function form that turn out to be extensions of the Deegan–Packel power index (Deegan and Packel 1978) and the Public Good power index (Holler 1982) defined on the family of simple games.

In the case considered and under the adoption of the plurality rule, the two new power indices proposed turn out to be sensitive to the different relevant ingredients present, such as the resulting minimal winning embedded coalitions and the number of representatives of each party or political grouping. They are also capable of reacting to the measurement of power in a changing and complex situation such as the National Assembly of Ecuador. It is worthwhile to further investigate these new proposals of power indices for weighted majority games in partition function form in the future. Some directions are the study of their properties and the achievement of axiomatic characterization results, the exploration of combinatorial analysis techniques aimed at efficient computation, and their application in other real-life problems.

Acknowledgements We would like to acknowledge the valuable comments of an anonymous referee. This work is part of the R+D+I project grants MTM2017-87197-C3-2-P, MTM2017-87197-C3-3-P, PID2021-124030NB-C32 and PID2021-124030NB-C33, that were funded by MCIN/AEI/10.13039/501100011033/ and by "ERDF A way of making Europe"/EU. This research was also funded by Grupos de Referencia Competitiva ED431C-2020/03 and ED431C-2021/24 from the Consellería de Cultura, Educación e Universidades, Xunta de Galicia.

References

Alonso-Meijide, J. M., Álvarez-Mozos, M., & Fiestras-Janeiro, M. G. (2017). Power indices and minimal winning coalitions for simple games in partition function form. *Group Decision and Negotiation, 26,* 1231–1245.

Álvarez-Mozos, M., & Tejada, O. (2015). The Banzhaf value in the presence of externalities. *Social Choice and Welfare, 44,* 781–805.

Álvarez-Mozos, M., Ferreira, F., Alonso-Meijide, J., & Pinto, A. (2015). Characterizations of power indices based on null player free winning coalitions. *Optimization, 64,* 675–686.

Álvarez-Mozos, M., Alonso-Meijide, J. M., & Fiestras-Janeiro, M. G. (2017). On the externality-free Shapley-Shubik index. *Games and Economic Behavior, 105,* 148–154.

Arévalo-Iglesias, G., & Álvarez-Mozos, M. (2020). Power distribution in the Basque Parliament using games with externalities. *Theory and Decision, 89,* 157–178.

Armijos-Toro, L. M., Alonso-Meijide, J. M., & Mosquera, M. A. (2021). *Mergeable weighted majority games and characterizations of power indices*. University of Santiago de Compostela.

Banzhaf, J. F. (1964). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19,* 317–343.

Barua, R., Chakravarty, S. R., & Roy, S. (2005). Weighted voting doesn't work: A mathematical analysis. *Homo Oeconomicus, 22,* 459–486.

Bolger, E. M. (1983). The Banzhaf index for multicandidate presidential elections. *SIAM Journal of Algebraic Discrete Methods, 4,* 422–458.

Bolger, E. M. (1990). A characterization of an extension of the Banzhaf value for multicandidate voting games. *SIAM Journal of Discrete Mathematics, 3,* 466–477.

Carreras, F., & Magaña, A. (2008). The Shapley-Shubik index for simple games with multiple alternatives. *Annals of Operations Research, 158,* 81–97.

Colomer, J. M., & Martínez, F. (1995). The paradox of coalition trading. *Journal of Theoretical Politics, 7,* 41–63.

de Clippel, G., & Serrano, R. (2008). Marginal contributions and externalities in the value. *Econometrica, 76,* 1413–1436.

Deegan, J., & Packel, E. W. (1978). A new index of power for simple n-person games. *International Journal of Game Theory, 7,* 113–123.

Dutta, B., Ehlers, L., & Kar, A. (2010). Externalities, potential, value and consistency. *Journal of Economic Theory, 145,* 2380–2411.

Holler, M. J. (1982). Forming coalitions and measuring voting power. *Political Studies, 30,* 262–271.

Holler, M. J., & Packel, E. W. (1983). Power, luck and the right index. *Journal of Economics, 43,* 21–29.

Johnston, R. J. (1978). On the measurement of power: Some reactions to Laver. *Environment and Planning A, 10,* 907–914.

Myerson, R. B. (1977). Values of games in partition function form. *International Journal of Game Theory, 6,* 23–31.

Shapley, L. S. (1953). A value for n-person games. In A. W. Tucker (Ed.), *Contributions to the Theory of Games II* (pp. 307–317). Princeton University Press.

Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48,* 787–792.

Thrall, R. M., & Lucas, W. F. (1963). n-Person games in partition function form. *Naval Research Logistics Quarterly, 10,* 281–298.

van den Brink, R., Dimitrov, D., & Rusinowska, A. (2021). Winning coalitions in plurality voting democracies. *Social Choice and Welfare, 56,* 509–330.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Measuring Voting Power in Complex Shareholding Structures: A Public Good Index Approach

Izabella Stach, Jacek Mercik, and Cesarino Bertini

1 Introduction

A complex corporate shareholding network consists of direct and indirect ownership relations. For example, company A has a share in company B, company B has a share in company C, company C has a share in company D, and company D has a share in company B. Is there some relation between firm A and company D? One can ask about the control power of company A in company D. This question was posed by many scholars starting from the last century; see Berle and Means (1932) for example. Since then, many researchers have tried to propose methods for measuring such "indirect" control power of a firm in ownership structure. Although it is not our intention to review all the methods proposed so far, let us list a few of those with a cooperative game theory approach and in particular that used power indices to the measurement of the control power of firms in corporate networks: Gambarelli and Owen (1994), Turnovec (1999), Hu and Shapley (2003a, 2003b), Leech (2002), Crama and Leruth (2007, 2013), Karos and Peters (2015), Mercik and Lobos (2016), Levy and Szafarz (2017), Mercik and Stach (2018), Stach et al.(2020), Staudacher et al. (2021a, 2021b), and Stach and Mercik (2021). The applications and comparisons of some of these methods can be found in Bertini et al. (2016), Kołodziej and Stach (2016), Stach (2017), and Mercik and Stach (2018). Not all of the methods proposed can be applied to complex shareholding networks. In corporate

I. Stach (✉)
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: istach@agh.edu.pl

J. Mercik
WSB University in Wroclaw, Fabryczna 29/31, 53-609 Wroclaw, Poland
e-mail: jacek.mercik@wsb.wroclaw.pl

C. Bertini
Department of Economics, University of Bergamo, Bergamo, Italy
e-mail: cesarino.bertini@unibg.it

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_9

shareholding structures where circular cross-ownerships exist, many methods fail. Only a few of the methods considered in the literature measure the control power of all firms involved in corporate shareholding structures. Among the methods that deal with the problem of cycles and measuring the control power of all firms in the corporate shareholding networks is one proposed by Karos and Peters in (2015). In this paper, we focus on this method and try to modify it using Holler's (1982) approach to measure players' a priori voting power in simple games.

We propose a new game-theoretical method (which is based on Holler's (1982, 2018) and Karos and Peter's (2015) approaches) to measure a firm's indirect control power as elements of a whole corporate ownership network. We called our approach briefly the *iPGI* index, where "*i*" refers to "indirect control." Our method considers voting rights attached to individual firms (represented by nodes in networks) as input data, i.e., direct and indirect ownership relations. The Karos and Peters method introduced in 2015 is used to create a new method. More precisely, in the Karos and Peters framework, instead of the use of the Shapley and Shubik (1954) index, we propose to use the Public Good Index (PGI for brevity) introduced by Holler in (1982). In this way, we obtain a method to measure the control power of all firms involved in complex shareholding structures (which means investors—firms without shareholdings—and stock companies). Moreover, this method is used to estimate the indirect control power in a theoretical example of a shareholding structure and compared it with the Karos and Peters approach. Finally, we also try to critically discuss the appropriateness of using the PGI index to measure the control power of firms in complex corporate networks.

As mentioned above, in 1982 Manfred Holler proposed a power index based on minimal winning coalitions—the Public Good Index called the Holler index (Holler, 1982, 2018; Holler and Packel, 1983). Holler and Li (1995) generalized this index from simple to cooperative games. Generally speaking, the difference between the PGI index and the very known normalized Banzhaf (1965) index lies in those winning coalitions that are not minimal (Holler & Nurmi, 2013).

A minimal winning coalition includes only those firms needed to obtain a majority position, and a minimal winning coalition consists of the minimal number of firms that can form a majority and control the smallest possible majority of the seats in the boards of companies. Some scholars (Holler, (1982, 1998), Riker, (1986) and others) claim that only minimal winning coalitions are coalitions that can occur and remain stable. Adding superflux players to the minimal winning coalition is costless and time-consuming. "This does not mean that surplus coalition do not form, but they should not be considered when measuring power"—this is cite taken from Holler and Nurmi (2013). This argument seems pertinent in the context of coalition formation in corporate networks. In particular, when we consider the formation of a coalition to a possible takeover of a company.

This paper aims to apply the PGI index to measure the indirect control power. To the authors' knowledge, in Staudacher et al. (2021a, 2021b), the PGI index was first used to measure all firms' indirect control power in complex corporate structures with cycles. In that paper, Staudacher et al. (2021a, 2021b) introduced and analyzed a framework of so-called implicit power indices generalizing the implicit indices

introduced by Mercik and Łobos (2016) and then modified in (Mercik & Stach, 2018; Stach et al., 2020; and Stach & Mercik, 2021), by replacing the Johnston (1978) index with several other power indices. Among these indices was the PGI index (Holler, 1982). The implicit power index takes into account not only the power of the individual entities constituting the companies (investors) but also the impact of the companies themselves on implicit relationships. In this paper, we follow the idea of Staudacher et al. (2021a, 2021b) to use the PGI index in the context of indirect control in complex shareholding structures. However, our framework is different as we based it on the Karos and Peters (2015) approach and tried to highlight the pros and cons of the PGI index in this context.

In the literature on power measurement—see Felsenthal and Machover (1995, 1998, 2005), Holler and Owen (2000, 2002a, 2002b), Laruelle and Valenciano (2011), and (Bertini et al., 2013)—there is an ongoing debate on the concept of power in general. For example, the literature discusses what we measure when we apply power measures and what properties an adequate measure of power should satisfy. In the context of indirect control, where we consider a possible acquisition of a company o group of companies, the question about the formation of minimal winning coalitions seems very important.

The rest of the paper is organized as follows. Section 2 provides the necessary background on simple games and power indices. Section 3 focuses on the key issue, i.e., the definition of a measure of the power of firms in complex corporate shareholding structures. This measure, which we call the *iPGI* index, is a modification of the Karos and Peters (2015) approach and is based on the index proposed by Holler (1982)—the Public Good Index. In this section, we also briefly discuss some properties, among other things, that axiomatically characterized the Karos and Peters index, and we try to check which of these are possessed by the *iPGI*. Section 4 provides a theoretical example of corporate shareholding structure in which we illustrate how the *iPGI* index can be applied in complex shareholding networks. We also compare the results thus obtained with the Karos and Peters approach. Finally, conclusions and some further developments are presented in Sect. 5.

2 Preliminaries and Notation on Simple Games and Power Indices

A *simple game* is a pair (N, v) consisting of a non-empty and finite set of *players* $N = \{1, 2, ..., n\}$ and a binary-valued function $v : 2^N \rightarrow \{0, 1\}$ domained on the set of all subsets of N—2^N, satisfying the following condition: $v(\emptyset) = 0$, $v(N) = 1$, and $v(S) \leq v(T)$ for all $S \subseteq T \subseteq N$. Any subset $S \in 2^N$ is called a *coalition*, and N is called the *grand coalition*. If $v(S) = 1$, then S is called a *winning coalition*; otherwise, $(v(S) = 0)$ it is called a *losing coalition*. By W and W_i we denote the set of all winning coalitions and the set of all winning coalitions containing player i, respectively, in a simple game (N, v). A player i is called a *critical* player in a winning

coalition S if $v(S\backslash\{i\}) = 0$. The set of all winning coalitions in which player i is a critical player is denoted by $\eta_i(v) = \{S \in W_i : v(S\backslash\{i\}) = 0\}$. A winning coalition S is called a *minimal winning coalition* if $v(S\backslash\{i\}) = 0$ for each $i \in S$. This implies that each proper subset of a minimal winning coalition is losing. By W^m and W_i^m we denote the set of all minimal winning coalitions and the set of all minimal winning coalitions containing player i, respectively, in a simple game (N, v). A player i is called a *null* player if there exists no coalition $S \in 2^N$ in which i is critical.

A *weighted game* $[q; w_1, \ldots, w_n]$ is a simple game (N, v) consisting of a non-negative vector of the weights of players $(w_1, \ldots, w_n)$, $\sum_{i \in N} w_i = 1$, and a majority quota $q \sum_{i \in N} w_i \geq q \geq \sum_{i \in N} \frac{w_i}{2}$ such that $v(S) = 1$ if and only if $\sum_{i \in S} w_i > q$. The weighted majority games are often used to model the voting situations in the stock companies.

2.1 Power Indices

For the purpose of measuring a priori the voting power of players in simple games, various one-point solutions called power indices have been proposed by different scholars. Generally, a *power index* f is a function that assigns a unique real-valued vector $f(v) = (f_1(v), f_2(v), \ldots, f_n(v))$ to each simple game (N, v). A component $f_i(v)$ assesses the power of player i in a simple game (N, v) for each $i \in N$. One of the best-known and frequently applied power indices was introduced by Shapley and Shubik (1954). The Shapley and Shubik index for a simple game (N, v) and each $i \in N$ is defined as follows

$$\sigma_i(v) = \sum_{S \in \eta_i} \frac{(s-1)!(n-s)!}{n!} \tag{1}$$

where $s = |S|$ denotes the cardinality of S. For more information on σ, see (Shapley & Shubik, 1954) or (Stach, 2011), for example.

The Public Good Index (PGI), also called the Holler (1982, 2018) index, for a simple game (N, v), and each $i \in N$ is given as follows

$$h_i(v) = \frac{|W_i^m|}{\sum_{j \in N} |W_j^m|} \tag{2}$$

The PGI index is also known as the Holler–Packel index, thanks to an axiomatization by Holler and Packel (1983) and then completed by Napel (1999).

2.2 Some Properties of Power Indices in Simple Games

In this section, we provide definitions of some desirable postulates of power indices in simple games. In particular, we quote only: bloc, efficiency, local monotonicity, null player, null player removable property, and symmetry properties. In Sect. 2.3, we will compare the power indices defined in Sect. 2.1 (σ and h) by taking these properties into consideration. The reason to discuss these properties and the σ and h power indices is to prepare a background to compare Φ and *iPGI*. As in Sect. 3.2, any difference in satisfying the equivalent properties in corporate networks observed in Φ and *iPGI* is the result of a difference in σ and h.

It is said that a power index f satisfies the following:

- The *bloc* property if, for all games (N', v') arising from the weighted game (N, v), $v = [q; w_1, ..., w_n]$ by removing two players $i, j \in N$ and introducing a new player representing the bloc $i\&j$ with $w_{i\&j} = w_i + w_j$, the following inequality holds $f_{i\&j}(v') \geq f_i(v)$;
- The *efficiency property* if, for all simple games (N, v), $\sum_{i \in N} f_i(v) = 1$;
- The *local monotonicity* (*dominance*) *property* if, for all weighted games $[q; w_1, ..., w_n]$ and any two distinct players $i, j \in N$, inequality $w_i \geq w_j$ implies $f_i(v) \geq f_j(v)$;
- The *null player property* if $f_i(v) = 0$ for all simple games (N, v) and each null player $i \in N$;
- The *null player removable property* if, for all simple games (N', v') arising from (N, v) by eliminating the null players, $f_i(v') = f_i(v)$ holds for each non-null player $i \in N$ ($i \in N'$ and $i \notin N \backslash N'$);
- The *symmetry property* if, for all simple games (N, v), each player $i \in N$, and every permutation $\pi : N \to N$, the following condition holds: $f_i(v) = f_{\pi(i)}(\pi(v))$, where $(\pi(v))(S) = v(\pi^{-1}(S))$;
- The *transfer property* if, for all pairs of simple games (N, v_1), (N, v_2) and each player $i \in N$, the following equation $f_i(v_1 \wedge v_2) + f_i(v_1 \vee v_2) = f_i(v_1) + f_i(v_2)$ holds, where $(v_1 \wedge v_2)$ and $(v_1 \vee v_2)$ are defined by the following sets of winning coalitions: $W(v_1 \wedge v_2) = \{S \in 2^N : S \in W(v_1) \text{ and } S \in W(v_2)\}$, $W(v_1 \vee v_2) = \{S \in 2^N : S \in W(v_1) \text{ or } S \in W(v_2)\}$.

2.3 Comparison of the Shapley and Shubik and PGI Indices

Let us consider an example of the voting system [51; 35, 21, 14, 15, 15] to compare the Shapley and Shubik index with the PGI index. In this system, there are five players, and the approval of a decision requires at least 51 votes of the total 100. The corresponding weights of players are: thirty-five votes for player 1, twenty-one votes for player 2, fourteen votes for player 3, and 15 votes each for players 4 and 5.

For each player i ($i = 1, 2, \ldots, 5$), Table 1 presents the set of minimal winning coalitions with player i, the set of winning coalitions in which player i is a critical player (η_i), and the distributions of power calculated by the Shapley and Shubik and PGI indices.

Table 1 Coalitions with critical players and distribution of power in the game [51; 35, 21, 14, 15, 15]

Player	W_i^m(set of minimal winning coalitions with player i critical)	η_i(set of winning coalitions with player i critical)	σ	h
1	{1, 2}, {1, 3, 4}, {1, 3, 5}, {1, 4, 5}	$W_1^m \cup \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 3, 4, 5\}\}$	48/120	4/14
2	{1, 2}, {2, 4, 5}	$W_2^m \cup \{\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{2, 3, 4, 5\}\}$	28/120	2/14
3	{1, 3, 4}, {1, 3, 5}	W_3^m	8/120	2/14
4	{1, 3, 4}, {1, 4, 5}, {2, 4, 5}	$W_4^m \cup \{2, 3, 4, 5\}$	18/120	3/14
5	{1, 3, 5}, {1, 4, 5}, {2, 4, 5}	$W_5^m \cup \{2, 3, 4, 5\}$	18/120	3/14

In this game set of the minimal winning coalitions consists of five coalitions ($W^m = \{\{1,2\},\{1,3,4\},\{1,3,5\},\{1,4,5\},\{2,4,5\}\}$). So, $|W_1^m| = 4, |W_2^m| = |W_3^m| = 2$, $|W_4^m| = |W_5^m| = 3$, and $\sum_{j\in\{1,\ldots,5\}} |W_j^m| = 14$. Then, the distribution of power by the PGI index immediately follows from formula (2), see Table 1.

Both indices are based on the concept of criticality of the player. However, the Shapley and Shubik index takes into account also all $n!$ possible orders to form the grand coalition. Therefore, in the formula of σ, *we* have the coefficient based on the cardinality of a coalition in which player i is critical. The PGI index considers only the number of minimal winning coalitions with player i. How a grand coalition is formed is not important in the PGI index. Player 2 is critical in one winning coalition of cardinality two, four winning coalitions of cardinality three, and one coalition of cardinality of size four. Thus, from formula (1), we have $\sigma_2(v) = \frac{3!}{5!} + 4 \cdot \frac{2!2!}{5!} + \frac{3!}{5!} = \frac{28}{120}$. Player 3 is critical only in two coalitions of cardinality three, so $\sigma_3(v) = 2 \cdot \frac{2!2!}{5!} = \frac{8}{120}$. Similarly we can calculate the power of players 4 and 5 by the Shapley and Shubik index: $\sigma_4(v) = \sigma_5(v) = 3 \cdot \frac{2!2!}{5!} + \frac{3!}{5!} = \frac{18}{120}$. As σ satisfies the efficiency property (see Shapley and Shubik (1954) and Sect. 2.2), then the power of player 1 is equal to $\sigma_1(v) = 1 - \sum_{i=2}^{5} \sigma_i(v) = 1 - \frac{52}{120} = \frac{48}{120}$.

Player 1 belongs to four of five minimal winning coalitions. Moreover, player 1 is critical in ten winning coalitions $|\eta_1| = 10$, the greatest value in this game. Therefore, there is nothing surprising in the fact that he obtains the greatest power by both σ and h indices, see Table 1. Players 2 and 3 belong only to only two minimal winning coalitions. So, according to the PGI index, they receive equal and the least power (2/14). The Shapley and Shubik index assigns player 2 much more power than player 3. Player 2 is critical in six winning coalitions, whereas player 3 is critical in only two.

The Shapley and Shubik index satisfies all properties mentioned in Sect. 2.2, whereas the PGI index does not satisfy the bloc, local monotonicity, and transfer properties (see Felsenthal and Machover (1995, 1998), Freixas and Gambarelli (1997), and Bertini et al. (2013), for example.

A violation of the local monotonicity property can be also observed in the above example. In particular, player 2, with more votes than player 4 (and player 5), obtains less power according to the PGI index, see Table 1.

Regarding the bloc property, for example, Bertini et al., (2013) provided a description of a failure of this property by the PGH index in the seven-players weighted game [6; 4, 1, 1, 1, 1, 1, 1] when a bloc between two players with weights 1 is forced.

A failure of the transfer property for the PGI index can be observed in the following pair of games: $v_1 = [4;3, 1, 1, 1]$ and $v_2 = [4;2, 2, 1, 1]$, see Bertini et al., (2013).

3 Indices for Measurement Indirect Control

In Sect. 3.1, we present the method introduced by Karos and Peters (2015) for measuring the indirect control of firms in the corporate shareholding networks. Then, in Sect. 3.2, we propose a modification of the Karos and Peters approach using the Public Good Index (Holler, 1982).

3.1 Karos and Peters Approach

Karos and Peters (2015) provide a method for measuring the power control of all firms in a corporate network (investors and stock companies as well). Namely, Karos and Peters propose the Φ index modelling the indirect relations among firms by so-called invariant mutual structures.

Let N be a set of all firms involved in a corporate shareholding structure. The invariant mutual control structure C is a function that assigns to each coalition $S \in 2^N$ the set of all firms controlled by S, such that satisfies:

- $C(\emptyset) = \emptyset$,
- The monotonicity property (i.e., $C(S) \subseteq C(T)$) for all coalitions $S \subseteq T \subseteq N$), and
- The indirect control condition (i.e., $\forall R, S, T \in 2^N$ with $S \subseteq C(T)$ and $R \subseteq C(S \cup T)$ we have $R \subseteq C(T)$).

Let denote by $\overline{C}$ the set of all invariant mutual control structures based on N. For every invariant mutual control structure C in $\overline{C}$ there is defined a vector of simple games $v^C = (v_1^C, ..., v_n^C)$. Each v_k^C indicates who controls firm $k \in N$ for C, and $v_k^C(S) = 1$ if k is controlled by S; otherwise, $v_k^C(S) = 0$. Thus, for every firm $i \in N$ in a shareholding structure, there is a simple game whose winning coalitions are exactly those that control i. The Φ index is defined as follows:

$$\Phi_i(C) = \sum_{k \in N} \sigma_i(v_k^C) - v_i^C(N) \; for \; every \; i \in N \; and \; C \in \overline{C} \tag{3}$$

where σ is the Shapley and Shubik (1954) index. In particular, $v_i^C(N) = 0$ for each null firm i, i.e., a firm that does not belong to any winning coalition to exercise its control, and i is also not controlled by any coalition.

The Karos and Peters method to measure indirect control in corporate shareholding networks is an axiomatic approach. They started from the set of properties that, in their opinion, should characterize a good measure and, as a consequence, came up with an index defined by formula (1). Namely, the Φ index satisfies five axioms: null player, constant sum, anonymity, transfer, and controlled player.

In particular, Karos and Peters null player axiom states that the power of null players is equal to zero.

The second axiom—the constant sum property—states that the sum of all assigned powers is the same over $\overline{C}$. The first (null player axiom) and second axiom imply that this sum is equal to zero.

The third axiom—anonymity—states that the names of the players should not matter.

The fourth axiom—transfer property—states that for each player, the change in power when enlarging a mutual control structure X to X' should be equal to the change in power when enlarging a mutual control structure Y to Y', assuming that the same control relations are added going from X to X' as when going from Y to Y'. The name is not casual, as this axiom is related to the transfer axiom used by Dubey (1975) to characterize the Shapley value and the Shapley-Shubik index (Shapley, 1953; Shapley & Shubik, 1954).

The fifth axiom—controlled player axiom—states that if firm i is a "controlled player", it means., controlled by at least one coalition and, as a consequence, by grand coalition N, but does not control any firm, then the power of firm i is set at -1. Subsequently, if firm j is an uncontrolled player, it means, controlled by no coalition at all, but firms i and j exert the same marginal control with respect to any coalition, then their difference in power is set at 1, i.e., firm j obtains 1 more than firm i.

For a precise definition of the Φ index and its properties, see Karos and Peters (2015).

3.2 *Holler-Based Estimation of Firm's Control Power*

What happens if we change the Shapley and Shubik (1954) index in the definition of the Φ index given by formula (3)? In the Karos and Peters (2015) framework, instead of the Shapley-Shubik index, we propose to use the Public Good Index (Holler, 1982). In this way, we obtain a modification of the Karos and Peters index—Holler-based estimation—for measuring firms' power control in corporate structures. In particular, this new index, *iPGI*, is defined by the following formula:

$$iPGI_i(C) = \sum_{k \in N} h_i(v_k^C) - v_i^C(N) \; for \; every \; i \in N \; and \; C \in \overline{C} \qquad (4)$$

where C, $\overline{C}$, and $v_i^C(N)$ are defined as in Sect. 3.1.

Generally, formula (4) implies that each player i obtains the sum of all his Public Good Index values in the games in which he contributes to controlling the other players, minus the sum of all Public Good Index values of the other players in the game describing the control undergone by player i.

Of course, if we change the σ index with the h index in formula (3), we have to take into account that some of the properties satisfied by the Φ index will not be met by the new index—*iPGI*. The question is which of the properties will be preserved.

The PGI index, as the Shapley and Shubik index, satisfies the null player property in simple games, thus also the Karos and Peters null player property is satisfied by the *iPGI* index.

The PGI index, as the Shapley-Shubik index, satisfies the efficiency property in simple games, i.e., the sum of power assigned to all players equals 1. Thus, this property, together with the null player axiom, makes that the sum of all assigned powers is the same over $\overline{C}$ is equal to zero. So, the *iPGI* index satisfies the constant sum property.

The third axiom—anonymity—is also satisfied by the *iPGI* index as the PGI index satisfies this property in simple games.

It is difficult for the fourth axiom (transfer property) to be satisfied by the *iPGI* index as the PGI index does not satisfy this property in simple games, see (Bertini et al., 2013} or Sects. 2.2–2.3 for example.

The five axiom—controlled player axiom—is satisfied by the *iPGI* index by the construction of the formula (4), and that h satisfies the null player property.

In the context of indirect control, the null player removable property—which states that after removing null players from a simple game the power assigned to non-null players remains the same—is one of the desirable properties that was highlighted first in (Mercik & Stach, 2018) and next in (Staudacher et al., 2021a, 2021b). Still, (Staudacher et al., 2021a, 2021b) offers a bit more cautious "null investor removable property" for corporate shareholding networks with distinguishable investors and companies. Namely, let's cite here this property: "After removing the null investors, i.e., the investors whose voting rights cannot transform any losing coalition into a winning one, from a corporate shareholding network with distinguishable investors and companies, the non-null firm's measures of power should remain unchanged. Equivalently, the value of any firm in a corporate shareholding network is unchanged if the network is extended by adding a new null investor." It was just noted in (Staudacher et al., 2021a, 2021b) that the Φ index fulfils the null investor removable property. The PGI index satisfies the null player removable property in simple games, and as a consequence also the *iPGI* satisfies the null investor removable property for corporate shareholding networks with distinguishable investors and companies.

In a weighted game, we say that a power index satisfies local monotonicity if a firm that controls a large share of the total weight vote does not have less power than a firm with a smaller voting weight. Felsenthal and Machover (1998) state that any power index that does not satisfy the local monotonicity property is "pathological" and should be disqualified as a valid measure of power. Holler and Napel in (2005) claim the following: "Power indices that detect rather than postulate monotonicity

can also be of help for a more abstract analysis of decision situations with respect to power." Holler and Napel in (2005), see the violation of this property as an advantage in some sense; see also Freixas and Kurz (2016). By the construction of formula (2), we see that any violation of the local monotonicity for the Public Good index (see Holler, 1982; Holler and Packel, 1983) implies a violation of the *iPGI* index. It is well-known that the Shapley and Shubik (1954) index satisfy the local monotonicity, see (Bertini et al., 2013), and also the index Φ possess this property. Moreover, the Φ index satisfies the monotonicity property proposed by Karos and Peters (2015) in the context of indirect control. Namely, their monotonicity postulate states that for two invariant mutual control structures X and Y: if a firm i is at least as much controlled in Y as he is in X, and his marginal control with respect to each coalition S is in X at least as large as in Y, then this firm should be assigned at least as much power in X as in Y.

The bloc property is also worth being mentioned, see (Mercik & Stach, 2018). The bloc property requires that the power of the merged entity $\{i\&j\}$—a bloc formed by player i with player j—will be larger than the power of player i if player j is not a null player. The bloc between players i and j, $\{i\&j\}$, may be regarded as a result of a takeover, in which player i, having annexed j*'s* voting rights, now trades under the new name $\{i\&j\}$. Intuitively, it seems reasonable that a player should not lose in annexing the voting rights of another player who is not null. The PGI index in simple games does not satisfy this property, see (Bertini et al., 2013). On the other hand, the measure that does not satisfy this postulate can be used to reveal this information. Let us cite Holler and Napel, (2005): "Obviously, (…) the bloc principle presuppose that votes are transferable, at least, to some extent. However, if vote transfers are voluntarily, then, in fact, we do not need these principles (in the form of axioms) because i will not form a bloc with j if the power of $\{i\&j\}$ is smaller than the power of $\{i\}$, unless i wants to give up power. But we need a measure that tells player i that he should not merge with player j in this case, what is a desirable property if a power measure can point out this 'dilemma'." The Banzhaf index (1965) violates the bloc postulate, but not the local monotonicity property. While the Shapley and Shubik (1954) index obeys both, as already mentioned in Sect. 2.

4 An Example

Let us consider an example of a corporate shareholding structure with 13 firms—five stock companies (Companies: 1, 2, 3, 4, 5) and eight investors, i.e., firms without shareholdings (Firms 6, 7, 8, 9, 10, 11, 12, 13), see Fig. 1.

Figure 1 shows direct and indirect ownership in a theoretical example of corporate shareholding structure already studied in (Stach, 2017; Mercik & Stach, 2018; and Stach & Mercik, 2021). Percentages of ownership are indicated next to the links (direct arrows). For example, Company 5 has 25 per cent of direct ownership (i.e., we regard this value—25% —as a percentage of own voting rights) in Company 2. Through Company 2, it also has an indirect ownership in Companies 1 and 3. This

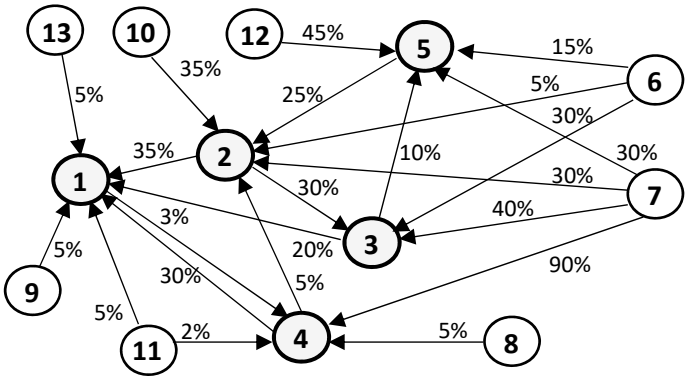


Fig. 1 Corporate shareholding network with 13 firms. *Source*: (Mercik & Stach, 2018)

theoretical ownership structure is not free of cycles, i.e., Company 5 owns 25% of Company 2, Company 2 owns 30% of Company 3, and Company 3 has 10% of Company 5, for example.

In this example, we regard a threshold of 50%—a simple majority. If a firm (or a coalition of firms) has ownership exceeding 50%, it has full control (100%), and the others have none (0%). In other words, with each stock company, we connect a weighted game with a simple majority.

Taking into consideration the direct and indirect ownership/control in this example, we can find the sets of all minimal winning coalitions for all companies (see Table 1), which facilitates to calculate the Φ and *iPGI* indices in this example (see Table 2).

In order to explain the result presented in Table 2, let's consider Company 1, for example. Taking into account only direct ownership, Company 2 with Company 3 have in total $35\% + 20\% = 55\%$ voting rights in Company 1, which gives coalition $\{2, 3\}$ full control over Company 1. Similarly, coalition $\{2, 4\}$ having 65% of voting rights can exert total control over Company 1. Next, Company 2 and Firm 6 have 60% voting rights in Company 3. Thus, coalition $\{2, 6\}$ controls Company 3. This implies that coalition $\{2, 6\}$ indirectly controls, via Company 3, Company 1. Similarly,

Table 2 Minimal winning coalitions in the example

Company	Minimal winning coalitions considering direct and indirect control
Co. 1	{2, 3}, {2, 4}, {2, 6}, {2, 7}, {3, 7, 9}, {3, 7, 11}, {3, 7, 13}, {5, 7}, {6, 7}, {7, 10}, {7, 12}, {3, 10, 12}, {3, 5, 10}, {3, 4, 9}, {3, 4, 11}, {3, 4, 13}, {4, 5, 10}, {5, 6, 10}, {6, 10, 12}
Co. 2	{5, 7}, {5, 10}, {6, 7}, {7, 10}, {7, 12}, {3, 10, 12}, {6, 10, 12}
Co. 3	{2, 6}, {2, 7}, {5, 7}, {6, 7}, {7, 10}, {7, 12}, {5, 6, 10}, {6, 10, 12}
Co. 4	{7}
Co. 5	{3, 12}, {6, 12}, {6, 7}, {7, 12}

considering indirect control, coalition {2, 7} controls Company 1. Firm 7, with 90% voting rights in Company 4, controls Company 4 totally. Thus, coalition {2, 7} controls Company 1. Then, coalition {3, 4, 9} has 55% voting rights in Company 1. As Firm 7 totally controls Company 4, then coalition {3, 7, 9} controls indirectly (via Company 4) Company 1 as well. Continuing consideration about direct and indirect ownership in Company 1, we have nineteen minimal winning coalitions that control Company 1.

Stach and Mercik in (2021) calculated the Φ index in this example. However, for clarity and to give the possibility to compare Φ with a new proposed index—*iPGI*, we present these calculations in Table 3.

The results in Table 3 are obtained by calculating the σ index for each company i in the weighted game of this company v_i, $i = 1, 2, \ldots, 13$. Let C be an invariant mutual control structure, based on $N = \{1, 2, \ldots 13\}$, defined by the vector of weighted games $(v_{11}^C, \cdots v_{13}^C)$. The set of all winning coalitions of each game v_i^C $(i \in N)$ indicates coalitions that control i. To calculate the σ index by formula (1) the set of all winning coalitions is needed, but this can be easy find when the set of all minimal winning coalitions in Table 2. For example, in Company 5, we have four non-null players (3, 6, 7, and 12) and four minimal winning coalitions: {3, 12}, {6, 12}, {6, 7}, {7, 12}, see Table 2. As the Shapley and Shubik index satisfies the null player property, then $\sigma_i(v_5^C) = 0$ for each player $i = 1, 2, 4, 5, 8, 9, 10$, and 11. The σ index satisfies the null player removable property as well, so we can calculate the power of non-null players considering a simple game with only these four players. The set of all winning coalitions consists of the four minimal winning

Table 3 Calculations of Φ index in the example

	Power distribution in accordance with σ index in simple game v_i							
Firm	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6, \ldots, 13$	Total	Φ
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–1.000
2	0.196	0.000	0.133	0.000	0.000	0.000	0.329	–0.671
3	0.121	0.017	0.000	0.000	0.083	0.000	0.221	–0.779
4	0.098	0.000	0.000	0.000	0.000	0.000	0.098	–0.902
5	0.056	0.150	0.050	0.000	0.000	0.000	0.256	–0.744
6	0.096	0.067	0.250	0.000	0.250	0.000	0.662	0.662
7	0.265	0.400	0.433	1.000	0.250	0.000	2.348	2.348
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.010	0.000	0.000	0.000	0.000	0.000	0.010	0.010
10	0.092	0.267	0.083	0.000	0.000	0.000	0.442	0.442
11	0.010	0.000	0.000	0.000	0.000	0.000	0.010	0.010
12	0.047	0.100	0.050	0.000	0.417	0.000	0.613	0.613
13	0.010	0.000	0.000	0.000	0.000	0.000	0.010	0.010
Total	1	1	1	1	1	0	5	0

coalitions and {3, 6, 7}, {3, 6, 12}, {3, 7, 12}, {6, 7, 12}, and {3, 6, 7, 12}. Player 3 is critical in only one two-person winning coalition {3, 12}, so from formula (1) we have $\sigma_3(v_5^C) = \frac{(2-1)!(4-2)!}{4!} = \frac{1}{12} \approx 0.083$. Both players 6 and 7 are critical in two two-person minimal winning coalitions and one three-person coalition. Thus, $\sigma_6(v_5^C) = \sigma_7(v_5^C) = 2 \cdot \frac{(2-1)!(4-2)!}{4!} + \frac{(3-1)!(4-3)!}{4!} = \frac{3}{12} = 0.25$. The σ index satisfies the efficiency property, so $\sigma_{12}(v_5^C) = 1 - \frac{1}{12} - 2 \cdot \frac{3}{12} = \frac{5}{12} \approx 0.417$, see Table 3.

Having σ for each simple game $(v_i^C, i = 1, \ldots, 13)$ the results, in Table 3, are immediately obtained by formula (3). Consider Company 1, for example. Company 1 is controlled by grand coalitions, so $v_1^C(N) = 1$. Moreover, Company 1 does not have voting rights in any of the other companies. Thus $\sigma_1(v_i^C) = 0$, for each $i = 1, \ldots, 13$. So, $\Phi_1 = 0 - 1 = -1$. Each investor is not controlled by any firm, so $v_i^C(N) = 0$ for $i = 6, 7\ldots, 13$. Investor 6 takes part in winning coalitions in Companies 1, 2, 3, and 5. Thus his voting power in these companies calculated by σ is greater than zero. By summing up these values and subtracting 0 we obtain $\Phi_6 = 0.096 + 0.067 + 0.250 + 0.250 - 0 = 0.662$.

In order to calculate the power control of each firm in the theoretical example in accordance with the *iPGI* index, it is necessary to calculate first the power distributions of the Public Good Index (the h index) in all companies, which is provided in Table 4.

The results in Table 4 are obtained by calculating the h index for each company i in the weighted game of this company $v_i, i = 1, 2, \ldots, 13$. As the PGI index satisfies both null player and null player removable properties, we can consider each game $v_i, i = 1, 2, \ldots, 13$ as a game that consists of only those players that are members of minimal winning coalitions, which makes them non-null players. Consider Company 5, for example. Game v_5 is a four-person game with four minimal winning coalitions, see Table 2. Player 3 is critical in minimal winning coalition {3, 12}; player 6 is critical in minimal winning coalitions {6, 7} and {6, 12}; player 7 is critical in {6, 7} and {7, 12}; and player 12 in {6, 12} and {7, 12}. Thus, from formula (2), we immediately have: $h_3(v_5) = \frac{1}{8} = 0.125$, $h_6(v_5) = h_7(v_5) = \frac{2}{8} = 0.25$, $h_{12}(v_5) = \frac{3}{8} = 0.375$. Having h for each simple game $(v_i, i = 1, \ldots, 13)$ the *iPGI* index is immediately obtained by formula (4), see Table 4.

The *iPGI* index of thirteen firms in the example is (–1, –0.8073, –0.6288, –0.8980, –0.6823, 0.6788, 1.9411, 0.0000, 0.0408, 0.5391, 0.0408, 0.7348, 0.0408), whereas the Φ index is equal to (–1, –0.671, –0.779, –0.902, –0.744, 0.662, 2.348, 0, 0.01, 0.442, 0.01, 0.613, 0.01).

In the considered example, we observe some similarities and differences between the *iPGI* and Karos-Peters (Φ) indices in assessing the power control of firms in the corporate shareholding network. Table 5 presents rankings of the stock companies and investors separately in accordance with both indices. Namely, considering the ranking of companies, the difference is only in Companies 2 and 3. The positions of the remaining companies estimated by Φ and *iPGI* are the same. The *iPGI* index gives more control power to Company 3 than to Company 5 and 2, classifying Company 3 in the first position. The Φ index classifies Company 2 in the first position, next

Table 4 Calculations of *iPGI* in the example

Firm	Power distribution in accordance with h index in company i						
	1	2	3	4	5	Total	*iPGI*
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	– 1.0000
2	0.0816	0.0000	0.1111	0.0000	0.0000	0.1927	– 0.8073
3	0.1837	0.0625	0.0000	0.0000	0.1250	0.3712	– 0.6288
4	0.1020	0.0000	0.0000	0.0000	0.0000	0.1020	– 0.8980
5	0.0816	0.1250	0.1111	0.0000	0.0000	0.3177	– 0.6823
6	0.0816	0.1250	0.2222	0.0000	0.2500	0.6788	0.6788
7	0.1633	0.2500	0.2778	1.0000	0.2500	1.9411	1.9411
8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0408	0.0000	0.0000	0.0000	0.0000	0.0408	0.0408
10	0.1224	0.2500	0.1667	0.0000	0.0000	0.5391	0.5391
11	0.0408	0.0000	0.0000	0.0000	0.0000	0.0408	0.0408
12	0.0612	0.1875	0.1111	0.0000	0.3750	0.7348	0.7348
13	0.0408	0.0000	0.0000	0.0000	0.0000	0.0408	0.0408
Total	1	1	1	1	1	5	0

Table 5 Rankings of firms by Φ and *iPGI* indices in the example

Position	Power distribution in accordance with			
	Φ for stock companies	Φ for investors	*iPGI* for stock companies	*iPGI* for investors
1	Company 2	Firm 7	Company 3	Firm 7
2	Company 5	Firm 6	Company 5	Firm 12
3	Company 3	Firm 12	Company 2	Firm 6
4	Company 4	Firm 10	Company 4	Firm 10
5	Company 1	Firms 9, 11, 13	Company 1	Firms 9, 11, 13
6		Firm 8		Firm 8

Company 5, and in the third position Company 3. Other companies have the same position in rankings of control power in accordance with both indices.

How can we explain this? The PGI index is based on the minimal winning coalitions, see Sect. 2.1 or Holler (1982). If we consider direct and indirect ownership and participation of a firm in all minimal winning coalitions presented in Table 2, we see that Company 3 participates in the greatest number of minimum winning coalitions of all firms involved in the example under consideration, see Table 6. Precisely, it takes part in 10 minimal winning coalitions, while Company 2 belongs to only four minimal winning coalitions and Company 5 belongs to only five minimal winning coalitions. Moreover, Company 3, through the minimal winning coalitions to which it belongs, controls three companies: Company 1, 2, and 5. While Company 2 controls

only two Companies: 1 and 3, and Company 5 controls three companies: 1, 2, and 3; see Tables 2 and 6. On the other hand, the Φ index is based on the Shapley and Shubik index, which in turn is based on the concept of a player's average "criticality" to all winning coalitions with him, see formula (3) and the formula (1) of the Shapley and Shubik index in Sect. 2.1. So, even Company 3 belongs to more minimal winning coalitions than Company 2 in the weighted game corresponding to Company 1; Company 2 belongs to more winning coalitions in which it is critical. Namely, the direct ownership of Company 2 in Company 1 (i.e., 35%) is greater than the direct ownership of Company 3 in Company 1 (i.e., 20%); see Fig. 1. Players 2 and 3 form a minimal winning coalition. However, with the enlargement of the coalition by players 4, 6, or 7, for example, player 3 is no longer a critical player, whereas player 2 is still critical. Thus, $|\eta_2| > |\eta_3|$ (the number of winning coalitions in which player 2 is critical is greater than the number of winning coalitions in which player 3 is critical) in Company 1. As a consequence, the power assigned to Company 2 is greater than the power assigned to Company 3 by the σ index in the simple game corresponding to Company 1. Then, the total power assigned to Company 2 in simple games corresponding to Companies 1 and 3 is greater than the power assigned to Company 3 in simple games corresponding to Companies 1, 2, and 5. Eventually, the Φ index classifies Company 2 in the higher post than Company 3.

When it comes to investors' ranking, the *iPGI* index gives more control power to Company 12 than to Company 6, as opposed to the Φ index. The power of Firm 12 in Co. 5 is really strong ($h_{12}(v_5) = 0.375$) and this decides about its total power in the whole network that is greater than the power of Firm 6 calculated by *iPGI*.

Table 6 Firm's control and participation in minimal winning coalitions in the example

Firm i	Number of minimal winning coalitions containing Firm i	Number of companies controlled by minimal winning coalitions with Firm i (companies' names)
1	0	0
2	4	2 (Companies 1 and 3)
3	10	3 (Companies 1, 2, and 5)
4	5	1 (Company 1)
5	5	3 (Companies 1, 2, and 3)
6	5	4 (Companies 1, 2, 3, and 5)
7	9	4 (Companies 1, 3, 4, and 5)
8	0	0
9	2	1 (Company 1)
10	7	3 (Companies 1, 2, and 3)
11	2	1 (Company 1)
12	5	4 (Companies 1, 2, 3, and 5)
13	2	1 (Company 1)

Both indices rank the control power of Company 7 as first and assign much more power to Firm 7 than to other investors. Namely, the *iPGI* index gives Company 7 48% of all the power assigned to investors, and the Φ index allocates even more: 57%; see Tables 3 and 4.

The Public Good Index (Holler, 1982) and the Shapley and Shubik (1954) index satisfy the null player property (it means that null players obtain zero power), so it is not strange that both *iPGI* and Φ indices classify Firm 8 on the last position with control power equal zero. Firm 8 does not belong to any minimal winning coalition (see Table 2), so Firm 8 obtains null power in each company (it means in the weighted game related to each company). So, the difference between both indices (*iPGI* and Φ) is only in the second and third position in investors' control power ranking, see Table 5.

5 Concluding Remarks

This paper has drawn up a discussion about the possible use of the PGI index (Holler, 1982) to measure the power of firms in corporate shareholding networks in terms of properties that this index possesses and fails. We proposed an approach based on the modification of the Karos and Peters (2015) method and the PGI index, see Sect. 3.

A justification for selecting the PGI index to assess player control power in complex corporate networks and using it in the framework proposed by Karos and Peters (2015) was that the index is based on minimal winning coalitions. In its non-normalized version, i.e., the raw measure, it counts the number of times that a player belongs to a minimal winning coalition. Therefore, the values assigned to companies according to the index may reflect the power of firms to form such coalitions. In the context of a possible takeover and the speed of companies' actions, it seems plausible to regard this kind of coalition at first. From this point of view, it is interesting to consider other indices based on minimal winning coalitions like the shift index–proposed by Alonso Meijide and Friexas (2010), and the Deegan and Packel (1978) index in the scheme proposed by Karos and Peters (2015).

Further development can refer to the application of the *iPGI* index to the estimation of company value in a complex market seen as a network of firms, see (Mercik et al., 2021), (Gładysz et al., 2019), and (Forlicz et al., 2018).

Acknowledgements This paper was funded under subvention funds for the AGH University of Science and Technology in Krakow, Poland; from WSB University in Wroclaw, Poland; and from the University of Bergamo, Italy.

References

Alonso Meijide, J. M., & Freixas, J. (2010). A new power index based on minimal winning coalitions without any surplus. *Decision Support Systems, 49*, 70–76. https://doi.org/10.1016/j.dss.2010.01.003

Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law RevIew, 19*, 317–343.

Berle, A., & Means, G. (1932). *The modern corporation and private property*. Commerce Clearing House.

Bertini, C., Freixas, J., Gambarelli, G., & Stach, I. (2013). Comparing power indices. *International Game Theory Review*, *15*(2), 1340004–1–1340004–19. https://doi.org/10.1142/S0219198913400045

Bertini, C., Mercik, J., & Stach, I. (2016). Indirect control and power. *Operations Research and Decisions, 26*(2), 7–30. https://doi.org/10.5277/ord160202

Crama, Y., & Leruth, L. (2007). Control and voting power in corporate networks: Concepts and computational aspects. *European Journal of Operational Research, 178*, 879–893. https://doi.org/10.1016/j.ejor.2006.02.020

Crama, Y., & Leruth, L. (2013). Power indices and the measurement of control in corporate structures. *International Game Theory Review, 15*(3), 1–15. https://doi.org/10.1142/S0219198913400173

Deegan, J., & Packel, E. W. (1978). A new index of power for simple n–person games. *International Journal of Game Theory, 7*, 113–123. https://doi.org/10.1007/BF01753239

Dubey, P. (1975). On the uniqueness of the shapley value. *International Journal of Game Theory, 4*, 131–139. https://doi.org/10.1007/BF01780630

Felsenthal, D. S., & Machover, M. (1995). Postulates and paradoxes of relative voting power - a critical reappraisal. *Theory and Decision, 38*, 195–229. https://doi.org/10.1007/BF01079500

Felsenthal, D. S., & Machover, M. (1998). *The measurement of voting power: theory and practice, problems and paradoxes*. London: Edward Elgar Publishers. https://EconPapers.repec.org/RePEc:elg:eebook:1489

Felsenthal, D. S., & Machover, M. (2005). Voting power measurement: a story of misreinvention. *Social Choice*, *25*, 485–506. https://www.jstor.org/stable/41106712

Forlicz, S., Mercik, J., Stach, I., & Ramsey, D. (2018). The Shapley value for multigraphs. In T. N. Nguyen, E. Pimenidis, Z. Khan, & B. Trawiński (Eds.), *Computational Collective Intelligence, LNAI, 10th International Conference, ICCCI 2018 Bristol, UK, September 5–7, 2018 Proceedings, Part II* (vol. 11056, pp. 213–221). Cham: Springer. https://doi.org/10.1007/978-3-319-98446-9_20

Freixas, J., & Gambarelli, G. (1997). Common internal properties among power indices. *Control and Cybernetics, 26*, 591–604.

Freixas, J., & Kurz, S. (2016). The cost of getting local monotonicity. *European Journal of Operational Research, 251*(2), 600–612. https://doi.org/10.1016/j.ejor.2015.11.030

Gambarelli, G., & Owen, G. (1994). Indirect control of corporations. *International Journal of Game Theory, 23*, 287–302. https://doi.org/10.1007/BF01242945

Gładysz B., Mercik J., & Stach I. (2019). Fuzzy shapley value-based solution for communication network. In N. Nguyen, R. Chbeir, E. Exposito, P. Aniorté, B. Trawiński (Eds.), *Computational Collective Intelligence. ICCCI 2019. Lecture Notes in Computer Science* (vol. 11683). Cham: Springer

Holler, M. (1982). Forming coalitions and measuring voting power. *Political Studies*, *30*, 262–271. https://doi.org/10.1111/j.1467-9248.1982.tb00537.x

Holler, M. J. (1998). Two stories one power index. *Journal of Theoretical Politics, 10*(2), 179–190. https://doi.org/10.1177/0951692898010002002

Holler, M. J. (2018). The Public Good index: A Brief Introduction (May 12, 2018) (vol. 1). Finnish-German Yearbook of Political Economy. https://ssrn.com/abstract=3598861

Holler, M. J., & Li, X. (1995). From public good index to public value. An axiomatic approach and generalization. *Control and Cybernetics, 24*, 257–270.
Holler, M. J., & Napel, S. (2005). Local monotonicity of power: Axiom or just a property? *Quality and Quantity, 38*(5), 637–647. https://doi.org/10.1007/s11135-005-4755-5
Holler, M. J, & Nurmi, H. (2013). Reflection on power, voting, and voting power. In M. J. Holler, & H. Nurmi (Eds.), *Power, Voting, and Voting Power: 30 years after* (pp. 1–24). Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-35929-3_1
Holler, M. J., & Owen, G. (Eds.) (2000). *Power measures*. Homo Oeconomicus XVII (1/2) (vol. 1). Münich: Accedo Verlag
Holler, M. J., & Owen, G. (Eds.) (2002a). *Power measures. Homo Oeconomicus*, XIX (3) (vol. 2). Münich: Accedo Verlag
Holler, M. J., & Owen, G. (2002b). On the present and future of power measure. In M. J. Holler, & G. Owen (Eds.), *Power Measures*, *Homo Oeconomicus* (vol. 19(3), 281–295). Münich: Accedo Verlag
Holler, M. J., & Packel, E. W. (1983). Power, luck and the right index. *Zeitschrift für Nationalökonomie (Journal of Economics)*, *43*, 21–29. https://www.jstor.org/stable/41798164
Hu, X., & Shapley, L. S. (2003a). On authority distributions in organizations: Controls. *Games and Economic Behavior, 45*, 153–170. https://doi.org/10.1016/S0899-8256(03)00023-X
Hu, X., & Shapley, L. S. (2003b). On authority distributions in organizations: Equilibrium. *Games and Economic Behavior, 45*, 132–152. https://doi.org/10.1016/S0899-8256(03)00130-1
Johnston, R. (1978). On the measurement of power: Some reactions to Laver. *Environment and Planning A, 10*(8), 907–914. https://doi.org/10.1068/a100907
Karos, D., & Peters, H. (2015). Indirect control and power in mutual control structures. *Games and Economic Behavior*, *92*(2015), 150–165. https://doi.org/10.1016/j.geb.2015.06.003
Kołodziej, M., & Stach, I. (2016). Control sharing analysis and simulation. In T. Sawik (Ed.), *Conference Proceedings: ICIL 2016: 13th International Conference on Industrial Logistics* (pp. 101–108). Zakopane, Poland, AGH University of Science and Technology.
Laruelle, A., & Valenciano, F. (2011). *Voting and collective decision-making*. Cambridge Books.
Leech, D. (2002). Shareholder voting power and ownership control of companies. *Homo Oeconomicus, 19*, 345–371. https://doi.org/10.1007/978-3-642-35929-3_25
Levy, M., & Szafarz, A. (2017). Cross-ownership: a device for management entrenchment?. *Review of Finance*, *21*(4), 1675–1699. https://doi.org./https://doi.org/10.1093/rof/rfw009
Mercik, J., & Łobos, K. (2016). Index of implicit power as a measure of reciprocal ownership. In N. T. Nguyen, R. Kowalczyk, & J. Mercik (Eds.), *Transactions on Computational Collective Intelligence XXIII. Lecture Notes in Computer Science* (vol. 9760, pp. 128–140). https://doi.org/10.1007/978-3-662-52886-0_8
Mercik J., & Stach I. (2018). On measurement of control in corporate structures. In N. T. Nguyen, R. Kowalczyk, J. Mercik, & A. Motylska-Kuźma (Eds.), *Transactions on Computational Collective Intelligence XXXI, Lecture Notes in Computer Science* (vol. 11290, pp. 64–79). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-58464-4_7
Mercik, J., Gładysz, B., Stach, I., & Staudacher, J. (2021). Shapley-based estimation of a company value – concept, algorithms and parameters. In Roszkowska E., Szopa M. (Eds.), *Entropy* (Volume 23, Issue 12, 1598. Special Issue: Decision Making, Classical and Quantum Optimization Methods). https://doi.org/10.3390/e23121598
Napel, S. (1999). The holler-packel axiomatization of the public good index completed. *Homo Oeconomicus, 15*, 513–520.
Riker, W. H. (1986). The first power index. *Social Choice and Welfare*, *3*(4), 293–295. https://www.jstor.org/stable/41105843
Shapley, L. S. (1953). A value for n-person games. In H.W. Kuhn, & A.W. Tucker (Eds.), *Contributions to the Theory of Games II*. Princeton: Princeton University Press
Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*(3), 787–792. https://doi.org/10.2307/1951053

Stach, I. (2011). Shapley-Shubik index. In K. Dowding (Ed.), *Encyclopedia of Power* (pp. 603–606). SAGE Publications.

Stach, I. (2017). Indirect control of corporations: Analysis and simulations. *Decision Making in Manufacturing and Services, 11*(1–2), 31–51. https://doi.org/10.7494/dmms.2017.11.1-2

Stach, I., & Mercik, J. (2021). Measurement of control power in corporate networks. *Operations Research and Decisions*, *31*(1), 97–121. https://doi.org/10.37190/ord210106

Stach, I., Mercik, J., & Bertini, C. (2020). Some propositions of approaches for measuring indirect control power of firms and mutual connections in corporate shareholding structures. In N. T. Nguyen, R. Kowalczyk, J. Mercik, A. Motylska-Kuźma (Eds.), *Transactions on Computational Collective Intelligence XXXV, LNCS* (vol. 12330, pp. 116–132). Heidelberg: Springer. https://doi.org/10.1007/978-3-662-62245-2

Staudacher, J., Olsson, L., & Stach, I. (2021a). Implicit power indices and software for measuring indirect control in corporate structures. In N. T. Nguyen, R. Kowalczyk, A. Motylska-Kuźma, J. Mercik (Eds.), *Transactions on Computational Collective Intelligence XXXVI, Lecture Notes in Computer Science* (vol. 13010, pp.). Verlag: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-64563-5_4

Staudacher, J., Kóczy, L. A., Stach, I. Filipp, J., Kramer, M., Noffke,T., Olsson, L., Pichler, J., Singer, T. (2021b). Computing power indices for weighted voting games via dynamic programming. *Operations Research and Decisions*, *31*(2), 123–145. https://doi.org/10.37190/ord210206

Turnovec, F. (1999). Privatization, ownership structure and transparency: How to measure the true involvement of the state. *European Journal of Political Economy, 15*, 605–618.

Voting On or About God

Mario Ferrero

By its very nature, religion is a difficult subject for theories of rational decision-making. One area that has been subjected to analysis in terms of political economy and social choice theory is the election of popes and the functioning of the papal autocracy (Colomer & McLean, 1998; see the survey in Ferrero, 2019). But the choice of god(s), the nature of God, and the doctrines regarding it–in a word, theology, which is supposedly the core business of religion–seem impervious to rational decision-making, whether individual or collective. In this study, we restrict ourselves to collective decision-making and examine a few historical instances in which such decision processes were apparently at work: theological disputations at a king's court, decisions on whether or not to adopt Christianity, and the working of the councils of the early Christian church. It will be seen that, remote as they are, these events can be interpreted in terms that are familiar enough to collective choice theory–hopefully to the benefit of both the said theory and our understanding of religious history.[1]

[1] Manfred Holler has spent a good part of his research career applying social choice theory and game theory to unlikely subjects, with results that were both entertaining and illuminating. To the best of my knowledge he has never addressed voting about divinity, so this study may be an apt, if marginal, contribution to his line of research.

M. Ferrero (✉)
University of Eastern Piedmont, Vercelli, Italy
e-mail: mario.ferrero@uniupo.it

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_10

1 Theological Disputations

1.1 The Khazar Conversion Debate

Theological disputations summoned by a king, for the purpose of choosing a religion for the state, are not uncommon in history. An example involved a debate between a Chinese Buddhist monk and an Indian Buddhist monk summoned to the court of King Trisong Detsen of Tibet around the year 793 CE, upon which the king proclaimed the Indian variant of Buddhism as the official religion (Stein, 1972, 65–68; Norbu & Turnbull, 1972, 177). Buddhism had already begun to spread in the country since the seventh century, supported by various kings, but it was opposed by the traditional, shamanistic Bon religion supported by the aristocracy. Trisong Detsen was determined to promote Buddhism anyway so in his decision the Bon option was out of question. Hence the choice was between two options.

More interesting are disputations in which the range of choice included three or more options. The clearest example is provided by the Khazars, the only nation that was officially converted to Judaism in post-Biblical times.[2] The Khazars were a semi-nomadic Turkic people that in the seventh century CE established a major trading empire in the steppes north of the Caucasus from the Black Sea to the Caspian Sea; its independence was terminated in the late tenth century by Kievan Rus', which incorporated its territories. Given its geopolitical situation, Khazaria was in control of the major trading routes from Asia to Europe and for centuries functioned as a buffer state between the Byzantine Empire and both the nomads of the northern steppes and the Arab Caliphate. Unsurprisingly, both empires tried to draw Khazaria to their side and the Arabs fought several, ultimately unsuccessful wars to conquer it. Also unsurprisingly for a commercial hub on the Silk Road, Khazar cities were multi-ethnic and multi-religious, with the original Turkic paganism coexisting peacefully with Muslim, Christian and Jewish merchants; moreover, Jews from Byzantium came to settle in Khazaria in several waves, driven by the persecutions and attempts at forced conversion periodically unleashed by the Byzantine emperors. At some point, variously estimated from the 730s to the early 800s, the Khazar royalty and elite converted to Judaism, and the kingdom was later universally known to both Muslim and Hebrew sources for its Judaism–though the extent to which the conversion seeped down among the common people is debated. Religious pluralism and toleration, however, seem to have continued to prevail throughout the existence of the kingdom.

[2] Golden (2007) provides a comprehensive discussion of the sources for the conversion and its historical context; Khazanov (1994) provides the larger context of the religious conversions of Eurasian steppe societies. These studies, however, do mention the disputation but do not describe it. We rely on the account of the disputation contained in the so-called Khazar correspondence (Wikipedia 2022), an exchange of letters in Hebrew that occurred around 960 CE between Hasdai ibn Shaprut, an influential Jewish scholar, adviser and foreign minister of the caliph of Al-Andalus in Cordoba, and Joseph, the Khagan (i.e. king) of Khazaria at the time. The disputation is noted and discussed in the influential, if controversial, study by Koestler (1976, Chap. 2).

Since both the adjoining empires were symbols of ancient civilizations and high culture, and both had a meddlesome interest in things Kazhar, it is easy to see that the pagan Khazar elite must have felt a bit like a country mouse squeezed between two town mice; so adopting a world religion–as in many other peripheral societies–was a way for them to upgrade their status among the nations. It is also easy to see, however, that choosing either Christianity or Islam would have compromised Khazar independence and upset the balance of regional power, drawing the Khazars too close to one of the great contenders. Therefore, modern scholars emphasize a strategic motive for the choice of Judaism: keeping a symmetric distance from the two great powers, and at the same time playing safe because the Jews, uniquely among the monotheists, did not have a national state and hence could not be suspected of having ulterior motives beyond offering to share their religion. While this interpretation makes perfect sense, according to some of the sources the decision was made by arranging for a disputation, and this is interesting because it was structured as the search for a Condorcet winner.

As the story is told by King Joseph in the Khazar correspondence (Koestler, 1976, Chap. 2; Wikipedia, 2022), a God-fearing king arose, named Bulan, who "expelled the sorcerers and idolaters from the land". Following this, an angel appeared to him in his dreams, exhorting him to worship the only true God and promising that in return He would bless and multiply the king's offspring and make his kingdom last till the end of the world–a story obviously inspired by the Covenant between the Lord and Abraham in Genesis, implying that King Joseph claimed for his people the status of a chosen people even though not descended from Abraham. Even though King Bulan was willing to submit to the angel's bidding, however, this still left his choice open between the three Abrahamic religions. Upon hearing the news, the kings of the Byzantines and the Arabs sent envoys with great presents as well as some of their wise men to convert the king to their religion. Being wise, however, Bulan sent for "a learned Israelite" and then put the three sages together to discuss their doctrines. The outcome of this disputation, described below, was that the king decided for the religion of Israel; he then proceeded to have himself, his court, and "all his people" circumcised (although the Arabic sources maintain that most of the commoners did not convert) and sent for Jewish sages who would teach him the Law and establish the observances. Based on both Jewish and Arab sources, all this probably happened about 740 CE. Two generations later, i.e. about 800 CE, one of Bulan's grandsons, King Obadiah, strengthened the rule of the Law according to tradition, built synagogues and schools, and brought in Jewish scholars who explained to him the Bible, the Mishna, and the Talmud, and the order of divine services. Based on these and other concurring sources, modern scholars agree that the Khazar conversion was a three-step process: first, a generic decision against idolatry and in favor of some monotheism; second, Bulan's conversion to the basics of Judaism to the exclusion of the other monotheisms–perhaps implying only a rudimentary form of Judaism, relying on the written Torah alone and excluding all rabbinical literature and the observances derived from it, such as that of the Karaite sect; and third, a religious revival and the implementation of full-fledged Rabbinic Judaism under King Obadiah.

Let us now focus on how the disputation was resolved and yielded the king's conversion to Judaism. Predictably, in the first stage each sage refuted the opponents' arguments so they could not agree. The king sent them away and then reconvened them after three days, asking them again to argue with one another and come up with a response as to which religion was the best; predictably, again the debate got nowhere. To get out of the deadlock, Bulan then resorted to a device: he asked the Christian cleric which of the religions of the Jews and the Muslims was nearer the truth, and he answered that the Israelite religion was better. Turning to the Muslim sage, Bulan likewise asked him which of the religions of the Jews and the Christians was nearer the truth, and he answered that the Israelite religion was better. Based on these answers, the king chose Judaism as both the Christian and the Muslim had admitted its superiority as second-best. In effect, Bulan had asked both the Christian and the Muslim for a complete ordering of their preferences. Assuming that the sages voted sincerely–which seems reasonable in this case–the procedure amounts to following the Condorcet method: Judaism turns out to be the Condorcet winner because it is the candidate that beats every other candidate in pairwise contests. Table 1 reports the rankings of the candidate religions expressed by the three voters, with the most preferred candidate being given the highest score (3), and assuming that each voter would most prefer his own religion. In King Joseph's account, the Jewish sage was not asked for his complete ordering, but this is inconsequential as his preferences between the other two religions could not have affected the result; so in the first column, the Jew's scores are given for both possible rankings. The last column reports the total scores achieved by each religion, which shows that in this case the Condorcet winner is also the winner of a Borda count.

It may be that behind the sages' votes lay a paramount theological concern: Israel's religion was the first monotheism in history and was recognized by all the others as the fountainhead of monotheism, whereas the Christian religion was a corrupted descendant of it in Muslim eyes, and vice versa; so the votes just acknowledged this. Or it may be that a political preoccupation was paramount: Judaism at the time was not a contender for worldly power whereas Christianity and Islam were the backbones of the two major imperial adversaries; so each sage might have reasoned, if my religion cannot carry the day, let the Jews have it rather than my direct foe. Either way, the result does not change: the Condorcet method makes second preferences pivotal and resolves an otherwise insoluble decision problem. So if we believe this account (scholars now credit King Joseph's letter as authentic), King Bulan practically invented the Condorcet and/or the Borda method some five centuries before

Table 1 The Khazar conversion decision

Voters / Religions	Jew	Christian	Muslim	Total
Judaism	3	2	2	7
Christianity	1 or 2	3	1	5 or 6
Islam	2 or 1	1	3	6 or 5

Ramon Llull and seven centuries before Nicholas of Cusa, who are generally recognized as the formal discoverers of these procedures–an interesting insight into the archaeology of social choice.

1.2 The Theological Debate at the Mongol Court

A more complex, if open-ended, example of theological disputation occurred at the Mongol court in 1254 and involved William of Rubruck, a Franciscan friar from the French Flanders who traveled as far as the capital of the Great Khan Möngke at Karakorum on a mission to convert the Mongols to Christianity. On his return, he wrote a detailed report in Latin for King Louis IX of France who had sponsored his mission (Jackson & Morgan, 1990), which is an invaluable early source of information about, among other things, the religions that thrived among the Mongols and other steppe peoples. These included traditional shamanism, Buddhism (William was the first western observer to describe it), Islam, and Christianity as represented by the Nestorian church, a sect declared heretical in the early fifth century and which achieved an extraordinary missionary expansion along the Silk Road and as far as China. Like other steppe societies before and after them, the Mongols were religiously tolerant and inclusive; William found representatives of all those religions all along his road and especially in and around the imperial court, where some high-placed officials, whether Mongols or members of other allied or subject peoples, were Muslims and others were Christians. The khans, however, may have had their own religious sympathies but were committed to the unity of their empire, and hence avoided being too closely identified with any one religion; rather, they were keen to have all the religious professionals pray for their well-being, thus securing to the state the greatest possible favor from Heaven. At their last interview, Möngke disclosed his mind to William, telling him that the Mongols believe that there is only one God, but "just as God has given the hand several fingers, so he has given mankind several paths. To you God has given the Scriptures and you Christians do not observe them", which William acknowledged, "whereas to us he has given soothsayers, and we do as they tell us and live in peace" (ibid, 236–237). After that William, to his regret, had no further occasion to expound the Christian faith to the khan. And he admits frankly that when the theological debate was over, no one present asked to become a Christian (ibid, 235). So his evangelizing mission ended in utter failure.

Even so, the khan did summon a theological disputation among all religious representatives present, even though it went nowhere–the khan and his advisers did not make any decision, as they probably never intended to. Following William's account in Chapter XXXIII (Jackson & Morgan, 1990, 229–235), all the groups were asked to hold a conference and, in view of that, to put down their doctrinal claims in writing so that the khan could make up his mind. The Nestorians wrote out a chronicle from the creation of the world to the Last Judgment, the last section of which contained "some faults" that William pointed out to them. William and his companions simply wrote down the Nicene Creed. Asked by William, the Nestorians

said they wanted to argue with the Muslims first but William advised against this, saying that in the debate they should seek alliance with the Muslims against the Buddhists because the former, like the Christians, believe in one God whereas the latter do not; the Nestorians agreed to this. Then William asked them whether they knew how idolatry had arisen in the world, which they did not know, so they asked William to explain these things first and then allow them to speak. William then proposed to rehearse: since the Buddhists deny the existence of God, he asked the Nestorians to prove that he does exist, but they showed to be incapable of proving anything and could only repeat what the Scriptures say, which was pointless since the Buddhists do not believe in them. So they agreed that William would be the first to confront the Buddhists while the Nestorians would step in only if he should be worsted.

On the day of the meeting a great crowd assembled, made up of supporters of all the contestants. The khan did not attend but sent three secretaries as umpires–a Christian, a Muslim, and a Buddhist, although apparently, the contestants themselves decided the agenda and the sequence of pairwise contests while the umpires presided over the fairness of the proceedings. So, to begin, William was confronted by a Buddhist from China, who asked him which question he wanted to debate first, the origin of the world or the fate of souls after death–this because, William tells us, those Buddhists had borrowed from "the Manichaean heresy" the belief in dualism between good and evil and in reincarnation. William countered that the discussion should instead start from the conception of God, and the umpires agreed. Then William stated the doctrine of the one and only God and his opponent replied that there is one supreme god in Heaven and countless many under him, just as there are many rulers on earth. William then laid out the doctrine of God as all-powerful, all-knowing, and all-good; his opponent denied that and asked, if your God is as you say, why has he created evil? At which William replied that the world is all good and it is not God who created evil, but managed to skirt the consequent logical question of where evil comes from and pushed the Buddhist back to his statement about a supreme god: was this god all-powerful? Reluctantly, the Buddhist answered that no god is all-powerful, at which the Muslims burst into laughter. Then William began to argue the unity of the divine essence and the Trinity but the Nestorians stopped him and sought to confront the Muslims. The latter, however, refused to argue and conceded that the Christians' religion and everything in the Gospels are true. At this point, the Nestorians had a long discussion with an old Uighur priest, expounding everything down to the coming of Christ in the Last Judgment and also explaining the Trinity to him and to the Muslims. "Everybody listened without challenging a single word" (ibid, 235). Thus ends William's rather elliptical account–a non-conclusion.

The foregoing account shows at various points that William judged the Nestorians' theology defective and their rhetorical skills inadequate. More generally, he took a dim view of the Nestorian priests (Jackson & Morgan, 1990, 163–164, 199). They are utterly ignorant; they have the Holy Scriptures in Syriac, a language they no longer understand, so they chant by rote, "and for this reason they are completely corrupt". They are drunkards and usurers, borrow customs from Islam, and not only permit but participate in sorcery and divination. A bishop visits them hardly once

in fifty years, on which occasion all the male children, even those in the cradle, are ordained priests. They engage in simony, levying fees for all the sacraments. They marry, and re-marry if widowed; hence they care for their wives and children, and consequently for making money, more than for spreading the faith. As a result, by their immorality and their greed, they rather alienate from the Christian religion the Mongols and the Buddhists, whose lives "are more blameless than their own". As Jackson and Morgan remark in their Introduction, this contrast highlights the "gulf between these eastern Christians, with their own tradition of eight centuries' standing, and Friar William, who represented the newly founded intellectual and evangelistic shock-troops of the Latin West" (ibid, 47). It is also clear from William's account that both the Muslims and the Buddhists saw William and the Nestorians as distinct groups. If so, we are entitled to analyze the disputation as a four-way affair, which reveals the complete preference orderings of four voters over four religions, here called Roman (church), Nestorian (church), Islam, and Buddhism. Even though no actual "winner" was proclaimed to crown the disputation, looking for an implicit Condorcet winner (and, subordinately, for a Borda winner) may be a useful exercise to perform once the preference orderings are laid out–one which may even suggest, albeit speculatively, a reason why no winner was in fact proclaimed.

The preference orderings that we can reconstruct from William's account are laid out in Table 2. Beginning with William, he is a Latin churchman with a missionary, not a political purpose. We have seen above that he is heavily critical of the Nestorians; nevertheless, they are Christians, albeit corrupt ones, so he would rank them next to the Roman church. As he himself tells the Nestorians, Islam comes third and Buddhism last. Turning to the Nestorians, there are two plausible versions of their ranking; in both Buddhism comes last as idolatry, but in the first version (N1) the Roman church comes before Islam while in the second (N2) the opposite is the case. The first version prioritizes the unity of Christianity above all else, while the second makes room for the nontrivial consideration that the Muslims, like the Nestorians, have long been acquainted with life at court and among the Mongols at large, so the two groups, albeit rival, must have developed some kind of mutual accommodation, whereas William was a moralizing outsider and the Roman church a foreign force with a potential for disrupting the other groups' easy life around the court (a life at which William looks askance). For the Muslims this last consideration would definitely take priority in their ranking, putting the Nestorian ahead of the Roman church and the latter, of course, ahead of Buddhism. Finally, the Buddhists would obviously rank the outsider (the Roman church) last, but they too have long since been around and must have developed a kind of live-and-let-live arrangement with the other two regulars around the court, and there is no knowing which of them should be felt closer. So we have two versions of the Buddhist ranking, with B1 ranking Nestorian above Islam (perhaps more naturally, since William repeatedly remarks that the Nestorians dabble in pagan practices but says nothing of the kind about the Muslims) and B2 the other way round.

Table 3 translates the preference rankings of Table 2 into numerical scores for each religion from each voter in decreasing order of preference, with 4 being the most preferred (the voter's own religion) and 1 being the least preferred. In keeping

Table 2 The disputation at the Mongol court: preference orderings

Ranking	William	Nestorian 1	Nestorian 2	Muslim	Buddhist 1	Buddhist 2
1st	Roman	Nestorian	Nestorian	Islam	Buddhism	Buddhism
2nd	Nestorian	Roman	Islam	Nestorian	Nestorian	Islam
3rd	Islam	Islam	Roman	Roman	Islam	Nestorian
Last	Buddhism	Buddhism	Buddhism	Buddhism	Roman	Roman

with the discussion above, there are two columns for the Nestorians (N1 and N2) as well as two for the Buddhists (B1 and B2). The last four columns give the row totals (T) for each religion by each pair of (Nestorian, Buddhist) rankings, with the first subscript denoting the Nestorian column and the second denoting the Buddhist column used. For each (N, B) pair, the Condorcet winner (if one exists) can be found by comparing row with row, i.e. comparing every religion with every other religion pairwise and determining which candidate wins each pairwise contest. Since we have an even number of voters (four), a tie is possible. A Condorcet winner is a candidate that beats or ties with every other candidate in pairwise contests; if a tie occurs, we have a "weak" Condorcet winner. For both the pairs (N1, B1) and (N2, B1) the Condorcet winner is Nestorianism, which beats all other religions 3:1; thus, as long as the Buddhist ranking is B1, one or the other version of Nestorian preferences makes no difference. Things change somewhat, however, if the Buddhist ranking is B2. Buddhism here is what is called a Condorcet loser, i.e. a candidate that is defeated by every other candidate in pairwise contests, irrespective of the (N, B) pair chosen. Even so, the switch from B1 to B2 reverses the Buddhist preference between Nestorianism and Islam and strengthens the latter. As a consequence, in the pair (N1, B2) Nestorianism becomes a weak Condorcet winner as it beats Roman and Buddhism but ties with Islam (we assume that this result is still superior to Islam's, which now beats only Buddhism but ties with both Nestorianism and Roman–an even "weaker" Condorcet winner). Finally, and remarkably, in the pair (N2, B2) Islam comes out as a weak Condorcet winner on a par with Nestorianism: they both beat Roman and Buddhism and tie with each other. The total scores reported in the last four columns allow computation of the Borda count for each preference pair, where the Borda winner's score is typed in boldface. As with the Khazar debate, the Borda method here produces the same results as the Condorcet method: under (N1, B1), (N2, B1), and (N1, B2) the single winner is Nestorianism (in the last preference pair, the weak Condorcet winner translates into a lower score for Nestorianism – 12 as against 13), while in the last column (N2, B2) Nestorianism and Islam tie at a score of 12.

We have justified the alternative orderings for Nestorians and Buddhists as due to our insufficiently detailed information about their true preferences, on the assumption that they all voted sincerely. However, the switch from B1 to B2 could also be interpreted as the Buddhists' intentionally "burying" Nestorian under Islam, i.e. insincerely lowering Nestorian's ranking for the purpose of favoring Islam. Similarly, the switch from N1 to N2 could be interpreted as the Nestorians' intentionally

Table 3 The disputation at the Mongol court: scores and Borda counts

Voters / Religions	W	N1	N2	M	B1	B2	T_{11}	T_{21}	T_{12}	T_{22}
Roman	4	3	2	2	1	1	10	9	10	9
Nestorian	3	4	4	3	3	2	**13**	**13**	**12**	**12**
Islam	2	2	3	4	2	3	10	11	11	**12**
Buddhism	1	1	1	1	4	4	7	7	7	7

Note. The last four columns give the total scores for each pair of (Nestorian, Buddhist) preferences. Thus e.g. T_{12} is the total score for each religion under the preference pair (N1, B2)

"burying" Roman under Islam, i.e. insincerely lowering Roman's ranking for the purpose of favoring Islam. If *both* these switches occur simultaneously, the result is to promote Islam to the position of weak Condorcet co-winner (and Borda co-winner) with Nestorianism. These "buryings" would have occurred through the differential emphasis and doggedness with which the contestants would have argued with each other–something we cannot determine since, unfortunately, William is our only source and he does not provide enough detail for us to use as evidence of their behavior in the debate. Why would these players have done it, since the Buddhists gain nothing and the Nestorians actually slide from single winner to co-winner as a result? Clearly there must have been some "side payment" occurring outside of the disputation and involving some other dimension of the inter-group relationships at Karakorum. We must leave it there as a conjecture that cannot be substantiated, but it suffices to suggest that the strategic dimensions and the bargaining games involved might have been tangled and unpredictable enough for the khan and his umpires to find it wiser to wash their hands of the entire controversy.

2 Voting Over the Adoption of a New Religion

In the ancient polytheistic world, the worship of new gods often spread as private cults without involving collective decisions, but in many cases, the introduction of foreign gods required a formal decision by the state when it entailed a public festival and a new temple financed by public funds. Thus in the democratic Greek city-states–particularly in Athens, the best documented case–a vote was taken in the citizens' assembly over a new public cult, while in Rome a vote was taken by the senate (Ferrero, 2022, Chap. 2). In these religious systems, however, a new god might compete for funding with the traditional gods but there was never an issue that it could or should replace all the other gods as the exclusive religion.

Things changed with the introduction of monotheism, which is theologically exclusive by definition. Exclusive theology, however, is not the same as exclusive religious policy. The three great Persian empires (those of the Achaemenians, the Parthians, and the Sasanians), in all of which Zoroastrianism was the state religion,

never forced conversion on their subject peoples; the Sasanians even favored the establishment and spread of the Nestorian church as a pawn of imperial rivalry with the Greek Orthodox Church. The Arab conquerors of the seventh and eighth centuries set up the *dhimmi* system, whereby the "peoples of the Book" (i.e. Christians of every stripe and Jews), who constituted the vast majority of their new subjects, were allowed to retain their religion and laws as long as they did not proselytize and paid a head tax to their Muslim rulers–a policy later taken over by the Ottomans as the *millet* system. As a consequence, conversion to Islam was a long drawn-out and basically voluntary process, which arguably goes a long way toward explaining why Islam has remained a unitary, noncompetitive religion to this day (Ferrero, 2018). In stark contrast, forced conversion was the official policy of the Christian Roman Empire and was then extended to all the European nations and polities successively brought under Christian political control in the course of the Middle Ages. The model of missionary action followed by the church everywhere was a top-down process, where the mission aimed to convert the king, his family, and his nobility, and then the rest of the people would be carried along by hook or crook–the opposite of the bottom-up process that had marked the rise of Christianity in the empire.

However, there are hints that some collective decisions took place among the ruling elite. The Germanic chiefs and kings, beginning with the famous conversion of King Clovis of the Franks around 500 CE, were worried that their men might not follow them along if they converted, so they often convened a council of their warrior aristocracy (Ferrero, 2022, Chap. 5.2; Fletcher, 1997, Chaps. 1, 4). While we have no information about the proceedings of most of these councils, we do have some interesting details about the conversion of Edwin, the king of the Anglo-Saxon kingdom of Northumbria, who took baptism with his chief men in 627 CE under the influence of his Christian wife and a Roman priest. The story, as recounted in the eighth century by the great English church historian, Bede, gives some account of the procedure and motives of the decision; we are interested exclusively in the former, not in the latter. Edwin convened a council of his advisors and noblemen where the Roman priest explained the tenets of Christianity. Then Coifi, the pagan high priest in charge of the gods' temple and cult at court, declared that having been devoted to the gods all his life had brought him no tangible benefits, so turning to the new god might indeed be worthwhile. After some other councilor spoke in favor of the new religion, the king agreed to embrace Christianity and Coifi himself proceeded to destroy the idols and the temple forthwith.

The reasoning underlying Coifi's stance was spelled out in more detail on the opposite side of the conversion issue, in a letter that Bishop Daniel of Winchester wrote to the apostle of the continental Saxons, St Boniface, in the eighth century (Fletcher, 1997, 251–252). He argued that the pagans must be brought around by persuasion, not by force. If the pagan gods, he went on, were really as mighty, beneficent, and just as claimed, they would not only reward their worshipers but also smite their foes; but then, why have they spared the Christians who are wiping them off the map and smashing their idols? Likewise, while the Christians own fertile lands awash with wine, oil, and all kinds of riches, the pagans with their gods are being pushed back and left to rule over cold and miserable lands in northerly countries. While

Coifi's statement, as reported, sounds like circular reasoning–"I'm abandoning the old gods because they are abandoning me"–Daniel's argument provides an extended context for the progress of the new religion and the retreat of the old one in which each individual group's decision is taking place.

The same context, and the same implicit argument for conversion, may help us understand the conversion of Iceland. In Iceland, a commonwealth of free farmers led by local priest-chiefs, there was no king to convert, so the decision to adopt Christianity was made by democratic procedure–with a twist (Byock, 2001, 297–301). All chiefs with their followers convened annually at a general assembly called *Althing* to decide about the law and the settlement of disputes. Some settlers arriving in the tenth century were Christians from the Viking colonies of the North Atlantic islands, and some Christian missionaries arrived from Norway at the urging of a Norwegian king who, however, had no formal authority over Iceland. So by the end of the century, the population had a mixed allegiance, with some people in between honoring both the Christian and the pagan gods. Matters came to a head at the *Althing* of the year 1000, where the two factions argued their respective cases and were, understandably, unable to reach a consensus decision, raising the specter of secession and violence. Then the Christian leader, in a manner typical of the Icelandic style of conflict resolution, asked the Speaker of the assembly–himself a pagan and a *godhi* (a priest of the pagan gods)–to arbitrate the dispute. The Speaker retired to ponder matters for an entire night sheltering "under his cloak"–possibly consulting his gods for an answer. The following morning, to the disappointment of his pagan faction, he pronounced that henceforth everyone would be Christian and must take baptism. To help people swallow his decision, he made some concessions, including that the exposure of unwanted babies continued to be permitted, and that sacrifice to the pagan gods was still allowed as long as it was done in private–but of course, once deprived of its public face, paganism's days were numbered. It seems a fair guess that, faced with two parties of approximately equal strength, the Speaker under his cloak just decided to follow the tide–more and more peoples are converting around the world, so how can we hold back forever?

Scanty as our information about these proceedings is, there may be a way of rationalizing our small sample of stories in social choice terms. Coifi's, Bishop Daniel's, and (conjecturally) Iceland's Speaker's reliance on the general progress of Christianization around them makes sense if each of them sees his individual decision as embedded in a slow-motion, collective decision about adoption of Christianity by a committee that stretches across the relevant space (here, Europe) and across time, and whose members increase with time as ever more countries join the decision chain. Each successive decision-maker looks back to the decisions taken by previous decision-makers in the sequence; he believes that the previous voters are no fools but, presumably, smarter than a decision made by tossing a coin would be, and thinks of himself as just as smart as the others were–i.e. as having the same, higher-than-50% probability of making the correct decision. If these assumptions are reasonable, then our "voters" are, unknowingly but correctly, applying Condorcet's jury theorem: if the voters' independent probabilities of being right are the same and greater than those of being wrong, the probability of the committee reaching a correct decision

increases with the size of the committee. Then it makes sense for an individual voter to look back and consider the number of countries that have already converted: the greater this number, the higher the probability that the voter will be making the right decision by joining them. Of course, there is no objective truth in this particular decision, but this way of thinking must have relieved those decision-makers of the anxiety inherent in a dilemma that, on "objective" grounds, was impossible to resolve: one would have had to rely on pure faith precisely when the faith was not there, a contradiction. In the circumstances, reliance on a worldwide Condorcet jury might well have been the most rational way to save the day.

3 Voting About God in Early Church Councils

As is well known, beginning with the council of Nicaea in 325 CE, the newly legalized Christian church agonized for centuries over arcane theological issues in a series of councils which defined the mainline theology and consequently identified the losers, who were branded as heretics and often due for a grim fate. Most historians, however, have until recently focused on the leading figures of the debates, on their theology and their politics. Taking a novel approach, MacMullen (2006) read the *acta*, or minutes of the meetings, that have survived from all the church councils whose date and site can be located–some 250 for the two and a quarter centuries between Nicaea and the council of Constantinople of 553 – with a focus on the mass of ordinary bishops in attendance and on how the decisions were arrived at. Most of these councils were local assemblies or synods of bishops, but 25 were general gatherings summoned or authorized by the emperors, and five of these went down in church history as the "ecumenical" councils that enjoy doctrinal authority–even though the ecumenical label, which should have meant empire-wide, was often belied by the unequal and selective attendance, the western church usually being scantily and poorly represented (ibid, 67–68). We now review the main lines of MacMullen's reconstruction of the workings of those 25 emperor-summoned councils with the aim of understanding, to the extent possible, how any decision at all could be made on such impossibly complex issues.

In these councils, the democratic element was real enough. In contrast to the secret ballot used by the Roman senate, the church of those days–born and grown originally as a Greek institution—always used open public voting, whether at councils, in episcopal elections, or in the election of abbots and abbesses of monasteries. At councils, bishops sometimes voted on issues or motions by raising and counting hands, like in Greek city assemblies, but more often by massed shouts–voting by acclamation. These votes did really count, embodying the power of the majority; but at the same time unanimity was sought and forced by threats if need be, especially in councils summoned by emperors. Emperors, starting from Constantine, wanted a united church that could lend divine legitimacy to their power, not one split between wrangling factions. Hence, standing up for the "wrong" cause meant, at best, loss of face, at worst, excommunication and exile; so most bishops in most councils

eventually signed up to the propositions endorsed by the majority. Nevertheless, minorities of recalcitrant bishops did show up and go down on record in councils of the earlier part of our period, for example, a minority of 10% at Nicaea (most of whom subsequently yielded to reason and signed up) or of 25% at Constantinople in 381 CE (who subsequently retired), while such minorities disappear from view in the later period (ibid, 99).

A "cognitive" element here enters the picture. How could it be that large majorities of large numbers of bishops in attendance (for the ecumenical councils, numbers of participants ranged from about 150 at Ephesus I and Constantinople I to some 350 at Chalcedon) could make up their minds and vote about theological intricacies of forbidding difficulty? MacMullen (ibid, 30–31) culled a sample of some 50 theological questions that were raised at these councils. These range from "Is Christ begotten the equal of God unbegotten?" to "Is God the Father before the Son's existence?", from "Is Christ a copy of the Father or an image?" to "Is God's substance increased or divided in begetting?", from "Is 'subject' (*hypostasis*) the same as 'subsistence' or 'person'?" to "Is there one *hypostasis* or are there three?", from "Is the Pneuma-Spirit the equal of Father and Son?" to "Is Mary the mother of God, or of Christ, or of Jesus?". A minority of the bishops were illiterate, even though they knew large sections of Scripture by heart. The majority, however, had enjoyed a better than average education, so they thought (often wrongly) that they could handle such arcane questions, and as a result the questions themselves multiplied over time down to the level of hair-splitting. Far from being academic disputes, however, the answers to such questions given at councils had the–often realized–potential of breaking off the church in different directions, as happened in the wake of Nicaea, Ephesus I, and Chalcedon among others. Such a threat could materialize especially because the councils' final reports, or synodal letters, were often read aloud before local congregations, involving masses of ordinary people in partisan struggles the merit of which they could hardly figure out. Sermons in the church were also widely used to stir up popular partisan excitement in the aftermath of councils.

This led to massive violence, the third element of the picture. MacMullen (ibid, 56) counts a staggering total of no less than 25,000 deaths over the period, for the most part not clergy but ordinary people, and almost all of them were victims not of legal executions but of extra-legal killings and street violence and the ensuing intervention of the army. This was accompanied by massive physical destruction of premises, generally by arson. The issues that were fought over in most cases were strictly theological, however hard this is to believe, and the violence often peaked at the election of bishops who were seen as the men of one or another theological faction. Hundreds of bishops were forced off their sees over the period (ibid, 59).

The upshot of this discussion is that theological debates in councils were deadly serious affairs, upon which ecclesiastical careers and even lives depended. In particular, the 25 emperor-summoned councils that are our focus were adversarial events in which withering accusations of blasphemy were freely thrown at opponents and passions were acted upon in the loudest voices to prevail. So, squeezed between the democratic power of numbers, imperial pressures for unity and conformity, intense theological passions, the impossibility of parsing abstruse questions in a manner

understandable to most minds, and the threat of violence, how could council participants ever arrive at any decision? First, voting blocs were organized: most suffragan bishops were happy to give their metropolitan bishops a blank check and shout in unison with them when required (ibid, 82–83). Second, as mentioned above, once it became clear who the winners were going to be, it was wise to establish one's name among the winners, which inflated the recorded majorities relative to the true ones. So the key to the final resolution was the initial gaining of a majority. This was clearly achieved long in advance behind closed doors by the imperial authorities who summoned some great councils, such as Chalcedon's in 451 CE (ibid, 94–95) or Constantinople's in 553 which was prepared and managed by Emperor Justinian himself (ibid, 111–112). Absent such official direction, however, fine theological distinctions and controversial arguments could only be brought before the general audience after a restricted specialist discussion had sorted them out into simplified phrases, at which point the council's leadership was able to manage them. "Deep matters should be left to deep thinkers, as most bishops were content to agree" (ibid, 101). One would think that theological matters should be decided upon on the basis of Scripture, in which the bishops were well versed. But the choice of Scripture alone as a foundation for belief had in the course of time become too dangerous, lest one should inadvertently find himself on the wrong side of a past dispute already settled at previous councils and suffer accordingly. Better to take the safer course of synodal certification: the appeal to interpretive authority as established by the "Fathers" (i.e. the winners of previous councils) as long as it was short and simple. Thus Nicaea was enthusiastically reaffirmed, and after that Ephesus I, and then Chalcedon, and so on and on: these represented the "democratic" sorting out of the best opinions by the best judges of the church–that is, sticking to orthodox continuity (ibid, 110–111).

Management, then, was the winning card (ibid, 98–99). The president, whose authority derived from the emperor or from established church procedures, could usually steer the proceedings toward the desired outcome: he drew up the agenda with attendant documents (often forged or doctored) and witnesses, could intervene at any time to bend a discussion in some chosen direction, could call on speakers known to be reliable, was free to change the subject at any awkward juncture in the discussion, and could frame a statement or question so as to dictate the council's response. So it was always for the president to define what was to be decided, whereas no proposals seem to have arisen from the floor of councils; the mass of bishops had only their shouts to rely on when they wished to oppose or change something. These shouts must have had formal authority since, according to the *acta*, they were so often reported to the higher authorities, and so presidents must have paid attention to them–which explains the dissident minorities reported above from many councils, while on other occasions prior planning and maneuvering of the numbers attending secured virtual unanimity of voters.

Looking at the above account from the point of view of social choice theory, there is no mystery to the historical fact of large, mostly incompetent assemblies deciding over issues about divinity that could defeat the best-trained philosophical minds, and which over time built up the theological dogma of mainline Christianity. In most of these debates, the alternative formulations that were or could be proposed as answers

to a given theological question were more than two; or–what comes down to the same thing–the current question was intertwined with other questions that had already been answered in previous councils, and answering the current question implied that these previous answers had now to be either upheld or rejected. Think only of the Christological issue that tore Nicaea apart, and whose resolution–the *homoousios* definition–reverberated through all successive debates. Even if we make the heroic assumption that each voter had a consistent and complete preference ordering among all the possible combinations of answers to all the questions involved, it is more than likely that no combination would defeat all the alternative combinations in pairwise contests–if only because certain things lie beyond the power of words to express. In other words, there would be no Condorcet winner and the majority outcome would depend on the sequence in which the alternatives are put to the vote. Alternatively, since here we have to do with sets of interconnected judgments of a "true or false" kind rather than preference orderings, it may be more accurate to see the problem as one of voting on a set of premises versus voting on the conclusion that is now under judgment: each premise commands a majority, but the majorities involved are distinct and do not translate into a majority on the conclusion that should logically follow from those premises. If so, we are faced with the "discursive paradox" in the aggregation of judgments discovered by List and Pettit (2002, 2004) which intersects, but does not entirely overlap, with the Condorcet paradox in the aggregation of preferences.

In either interpretation, a way out of the impasse is to violate one or another of the standard conditions of majority decision-making: specifically, that no one voter should be given special weight in the collective decision, and/or that majority rule should be used to decide on each of the interconnected questions. The first option amounts to implicitly appointing a dictator–naturally, the presiding person at the council–whose job it is to ensure that some collective "rationality" (as judged by the president himself) prevails in the end. The second option amounts to prioritizing some of the questions and letting the decisions on the other, subordinate questions be determined by the decisions on the prioritized questions, not by majority rule. We have seen both strategies at work in our councils: the president had extensive agenda-setting and question-framing powers, by which he could usually ensure his preferred outcome by structuring the voting sequence in a certain way; and the decisions taken by previous councils on questions that could be argued to be logical premises to the question currently being decided were put above scrutiny by emphasizing the binding authority of the Fathers–as we have seen, to challenge an established dogma was to invite one's deposition, or worse. These strategies did not always work, nevertheless. Emperor Justinian summoned and personally presided over a small council at Constantinople in 532 to heal the festering rift between Chalcedonians and Monophysites, with only six bishops on each side; despite the small numbers, the bishops quarreled all the time and the emperor finally dismissed them in despair. Twenty years later, however, with the ecumenical council of Constantinople of 553, Justinian finally had his way: by writing a theological treatise to prepare for it, threatening or actually enforcing the deposition of Monophysite bishops, inviting only a handful of unruly Egyptians, pre-determining the sequence of topics for debate, and stage-managing the intervention of speakers, the emperor secured a well-behave

assembly that duly complied with his wishes (MacMullen, 2006, 107–112). Thus, through management from above and fencing around past decisions with anathema, were sown the seeds of the doctrine and practice of infallibility of ecumenical councils that would define the Catholic Church to this day.

4 Conclusion

We have examined three types of collective decisions about theological matters as exemplified by historical cases: theological disputations among representatives of several religions, each trying to entice a sovereign to embrace his own faith; councils convened by several Germanic peoples to decide whether to switch from paganism to Christianity; and councils convened by the early Christian church to decide its official doctrine on complex, difficult questions about the Trinity, the nature of Christ, and more. The available surviving evidence is unfortunately too scanty and lacking in relevant detail to permit anything like firm conclusions, but it is sufficient to suggest that some important concepts and results in social choice theory can be used to shed some light on those remote events. The Condorcet winner and Borda winner, Condorcet's jury theorem, and the Condorcet paradox all can be seen to play a role in those decision processes, unbeknown to the participants in each case. If these modern, sophisticated voting models do turn out to be applicable, this is good news for both historians, who may thus avail themselves of new tools to help their understanding of events, and economists and social choice scholars, who may thus pride themselves on yet another example of relevance of their tools to the real world. In particular, choice in religious matters, a topic seemingly impervious to rational choice, may after all prove within reach of scientific analysis–something worthy of further research.

Acknowledgements The author is indebted to Guido Ortona for useful conversation about the Khazar episode.

References

Byock, J. L. (2001). *Viking Age Iceland*. Penguin Books.

Colomer, J. M., & McLean, I. (1998). Electing popes: Approval balloting and qualified-majority rule. *Journal of Interdisciplinary History, 29*(1), 1–22.

Ferrero, M. (2018). Why the Arab spring turned Islamic: The political economy of Islam. *Constitutional Political Economy, 29*(2), 230–251.

Ferrero, M. (2019). Christian history and public choice. In R. Congleton, B. Grofman, & S. Voigt (Eds.), *The Oxford Handbook of Public Choice* (Vol. 2, pp. 760–780). Oxford University Press.

Ferrero, M. (2022). *The political economy of Indo-European polytheism: How to deal with too many gods*. Springer.

Fletcher, R. A. (1997). *The conversion of Europe*. HarperCollins.

Golden, P. B. (2007). The conversion of the Khazars to Judaism. In P. B. Golden, H. Ben-Shammai & A. Róna-Tas (Eds.), *The world of the Khazars: New perspectives. Selected papers from the Jerusalem 1999 International Khazar Colloquium* (pp. 123–162). Brill.

Jackson, P., & Morgan, D. (1990). *The mission of Friar William of Rubruck: His journey to the court of the Great Khan Möngke, 1253–1255*. Translated from the Latin by P. Jackson. Introduction, notes and appendices by P. Jackson and D. Morgan. London: Hakluyt Society.

Khazanov, A. M. (1994). The spread of world religions in medieval nomadic societies of the Eurasian steppes. In M. Gervers & W. Schlepp (Eds.), *Nomadic diplomacy, destruction and religion from the Pacific to the Adriatic* (vol. 1, pp. 11–33). Toronto: Toronto Studies in Central and Inner Asia.

Koestler, A. (1976). *The thirteenth tribe: The Khazar empire and its heritage*. Hutchinson.

List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy, 18*(1), 89–110.

List, C., & Pettit, P. (2004). Aggregating sets of judgments: Two impossibility results compared. *Synthese, 140*(1–2), 207–235.

MacMullen, R. (2006). *Voting about God in early church councils*. Yale University Press.

Norbu, T. J., & Turnbull, C. (1972). *Tibet: Its history, religion and people*. Penguin Books.

Stein, R. A. (1972). *Tibetan civilization*. Translated from the French by J.E. Stapleton Driver. Stanford University Press.

Wikipedia. (2022). Khazar correspondence. Retrieved February 22, 2022, from https://en.wikipedia.org/wiki/Khazar_Correspondence.

Probabilistic Study of Voting Rules: A Tale of Two Volumes

Hannu Nurmi

1 Introduction

I would like to begin with a brief personal note on my long scholarly cooperation and friendship with the honoree of this book, Manfred J. Holler. I have known Manfred for more than 40 years. What got us together was an interest in public choice and applied game theory, in those days not widely studied outside the United States. For economists, the public choice was apparently too 'political'. For political scientists, in turn, the game theory seemed too simplistic and the associated methodological individualism downright irrelevant. So, it took a hefty dose of professional courage for a young scholar to become an advocate and, indeed, a standard-bearer of public choice and applied game theory. And yet, Manfred soon became one. In the late 1970s, he founded *Munich Social Science Review* and a few years later *European Journal of Political Economy*, both journals of clear game theory and public choice emphasis. In the early 1980s, he edited one of the classics on applications of game theory to politics: *Voting, Power and Voting Power* Holler (1982). The study of voting power has been—and continues to be—one of the central themes in his scholarship. As a contributor to these fora, I shared much of Manfred's enthusiasm in these areas. He devised the public good index around the same time Holler (1978). Over the past decades, it has been widely and intensively debated and applied, most recently by Manfred and one of the editors of this volume Holler and Rupp (2020). I am mentioning these early works of Manfred because we seemed to be tiptoeing around voting procedures and probability models without actually combining the two as has been done in the two volumes discussed in this paper; one of these volumes appeared only a few years before our cooperation began. Truly grateful for the past decades

The author thanks Benoît Le Maux for constructive comments on an earlier version of this paper.

H. Nurmi (✉)
Department of Philosophy, Contemporary History and Political Science, University of Turku, Turku, Finland
e-mail: hnurmi@utu.fi

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_11

of friendship and cooperation, I very much hope that the following pages will be of interest to the honoree of this book.[1]

Voting is about making a choice, either of a person, of a group of persons or of a policy. No doubt choices are made in other ways as well, e.g. by bargaining, by lot, by delegation or by rules of succession. Yet, voting is often regarded as the most democratic way of making collective choices. Indeed, voting is typically regarded as a necessary, albeit not sufficient, condition of democratic governance. Over centuries, a large number of different voting systems have been devised for apparently the same purpose: to tease out the will of the voters. Since the systems are different, it follows that they at least occasionally result in different outcomes when processing the same voter input. This observation motivates the study of properties of various voting systems. At the same time, the question arises as to what constitutes a plausible or reasonable collective choice under various constellations of voter opinions.

This paper discusses a particular, viz. probabilistic, approach to the study of voting rules. It complements the evaluations that focus on norms, conditions or criteria that may or may not characterize the rules. The best-known results of social choice theory establish compatibilities or incompatibilities of those characteristics. Arrow's (1963) impossibility theorem on preference aggregation norms, Sen's (1970) impossibility result on Paretian liberal in resource allocation, Moulin's (1988) result on the incompatibility of participation and Condorcet's principle, as well as Gibbard's (1973) and Satterthwaite's (1975) theorem on general manipulability of social decision functions, are examples from this genre.

Alongside these relatively high-profile works, another approach to voting rules has developed, viz. the probabilistic modeling of voting systems. The main focus of these models is on the probability of encountering various kinds of anomalies and other peculiarities that are related to voting rules under specific circumstances. How likely are we to encounter such circumstances in practice? What kinds of factors contribute to the emergence of such circumstances? How different in the end are various voting rules in terms of the ensuing outcomes? In fact, probabilistic notions were present already in Condorcet's *magnum opus* Black (1958); Condorcet (1785). It must be said, though, that the early reception of these ideas was not uniformly enthusiastic. To wit, some 80 years after its publication, the probability aspect of Condorcet's work received a scorching criticism from Isaac Todhunter (Todhunter, 1865, p. 352):

> We must state at once that Condorcet's work is excessively difficult; the difficulty does not lie in the mathematical investigations, but in the expressions which are employed to introduce these investigations and to state their results: it is in many cases almost impossible to discover what Condorcet means to say. The obscurity and self contradiction are without any parallel, so far as our experience of mathematical works extends; some examples will be given in the course of our analysis, but no amount of examples can convey an adequate impression of the extent of these evils. We believe that the work has been very little studied, for we have not observed any recognition of the repulsive peculiarities by which it is so undesirably distinguished.

[1] Of course, this very brief exposition is not intended as a listing of Manfred's scholarly contributions, not even the most influential ones, but as a background to the following pages.

Indeed, had it not been for Nanson's and Black's rediscoveries, Condorcet's work on voting might well have landed in oblivion Black (1958); Nanson (1883). However, Condorcet's contributions that today bear his name are largely independent of his probability calculus. As the evolution of social choice theory in general, the probabilistic tradition has been highly discontinuous. The first book-length work in the present authors' knowledge did not appear until 1970s. The recent (2021) publication of *Evaluating Voting Systems with Probability Models. Essays by and in Honor of William Gehrlein and Dominique Lepelley* edited by Diss and Merlin (2021) nearly coincides with the 50'th anniversary of the pioneering volume in this genre: *Probability Models of Collective Decision Making* edited by Niemi and Weisberg (1972). This paper aims to provide a brief and non-technical overview of the development of probabilistic voting research in the light of the two volumes just mentioned.

Unsurprisingly, no author has contributed to both of these books.[2] What the volumes have in common is that they deal with probabilistic methods in addressing problems related to voting. A glance at the respective tables of content reveals similarities and differences in foci: constitutional design, coalition formation, and spatial models of party competition are predominant topics in the Niemi and Weisberg book, while the distinctive subjects of the slightly more voluminous Diss and Merlin volume are voting paradoxes and manipulability of voting rules. The study of cyclic majorities has an important place in both volumes. It is only natural and, at the same time, a measure of progress that the Diss and Merlin volume gives a more nuanced picture of the paradox of cyclic majorities and the related concept, Condorcet efficiency, i.e. the probability of a Condorcet winning alternative being elected by various voting methods. The progress of the literature on probabilistic modeling is also reflected in the set of chapters of the Diss and Merlin volume focusing on computational techniques, a subject not much known outside computer science in early 1970s.

In a way, probabilistic modeling represents a return to the roots of voting theory since in the late eighteenth century Marquis de Condorcet was preoccupied with jury decision-making under the assumption that the jurors have an individual probability of passing a correct judgment in dichotomous decision settings Black (1958). The task the Marquis set for himself was to find a decision rule that would maximize the probability of the collectivity making correct decisions. His main result—nowadays known as the Condorcet Jury Theorem (CJT)—suggested that if all voters have the same probability of making the correct choice in dichotomous choice situations and if this probability is higher than $1/2$, then the probability of the majority being right in these kinds of choice situations is larger than the individual probability and tends to unity with the increase of the number of voters. In contrast, if the individual probability of being right is smaller than $1/2$, the probability that the majority is correct is smaller than the individual one and tends to zero as the number of voters grows without limit. The CJT approach is still being pursued today, but is—

[2] While not an author in the Diss and Merlin volume, Peter C. Fishburn—one of the contributors to the Niemi and Weisberg book—certainly had an important role as the supervisor and mentor of William V. Gehrlein, a contributor to and one of the two honorees of Diss and Merlin (2021).

somewhat surprisingly—represented in neither the Diss and Merlin nor in the Niemi and Weisberg volume. Rather the starting point is the standard one of each individual being endowed with complete and transitive preference relations (rankings) over the alternatives. Probability enters the picture via assumptions about the distributions of voters over the preference of rankings.

2 The Principles of Probability Modeling of Voting Rules

As all scholarly activity, probabilistic modeling can be characterized by its goals and methods for achieving those goals. Probability models aim at studying the outcomes ensuing from various voting rules in specific environments. The outcomes depend on the preference profiles of candidates as well as on the way preferences are related to voting strategies. Very often the focus is on reported preferences, i.e. the voting strategies rather than on the underlying 'true' preferences. Thus, the issues related to sincere vs. sophisticated voting are glossed over. This is also the case in the contributions contained in the earlier of the two volumes discussed here. The study of outcomes resulting from voting rules may reveal incompatibilities between various choice desiderata or some intuitively bizarre occurrences. Some of the latter are known as paradoxes. The specific contribution of the probability models is to address the question of how common these occurrences are in various environments, i.e. in various types of preference profiles. It is, however, also possible to consider the likelihood of benefiting from sophisticated voting in various profile types. Regarding the role of probability models in the study of voting rules Gehrlein and Lepelley (2004, p. 141) conclude that:

> Generally speaking, impossibility theorems are essentially qualitative: they state that some paradoxes or difficulties are possibilities but leave open the likelihood of their occurrences. Thus, in the absence of empirical data, probability calculations have proved to be useful.

Another contribution of probability models in these authors' view is in the comparison of voting rules that in principle all fail on a specific criterion or desideratum. The models can then provide important information regarding their likelihood of failure in specific environments. The procedure is straightforward in principle. First, generate all profiles of the desired number of voters over the desired set of candidates. Second, tally the number of profiles where the procedure under study violates the desideratum of interest (e.g. yields a cyclic collective preference relation, elects a Condorcet loser or leads to a voting paradox of some sort). If all profiles are assumed to be equally probable, the tally divided by the number of possible profiles gives the probability of desideratum violations. The population of profiles from which the specific profile sets are drawn (e.g. those exhibiting majority cycles or single-peaked preferences) can simply be generated by sampling one ranking at a time or obtained by constructing probability representations. The latter is typically used in obtaining analytic formulae for relevant probabilities (e.g. of violating the Condorcet principle or choosing the Condorcet loser or ending up with different winner sets).

Two types of profiles stand out in the literature: (i) impartial culture (IC) profiles, and (ii) impartial anonymous culture (IAC) profiles. In the former, the individual preference rankings are randomly and independently drawn (with replacement) from the population of all possible rankings so that any conceivable ranking is equally probable. In the latter, all conceivable *distributions* over the preference relations are deemed equiprobable. Both cultures are intuitively pretty far from describing any real-world electorate, but (ii) is more commonly used. While both (i) and (ii) have very limited descriptive value, they provide a useful benchmark for comparative evaluations of voting rules in terms of various norms and criteria.

3 Common Themes in the Two Volumes

It is only natural that in an active research field topics of interest change with the passage of time. Hence, one can expect to find many new topics in Diss and Merlin (2021) that were not dealt with in Niemi and Weisberg (1972). Similarly, many themes focused in the latter have been omitted in the former volume.

3.1 Cyclic Majorities

One of the common themes in Niemi and Weisberg (1972) and Diss and Merlin (2021) pertains to concepts related to the name of Condorcet. However, the concept of Condorcet winner or loser was not yet in standard use in the early 1970s and not directly applied in the former volume, but the notion of majority cycle was well-known and a focus of interest in the Niemi and Weisberg volume. Bowen discusses the possibility of cyclic majority preference relation underlying the votes in the U.S. Senate (Bowen 1972). It is well-known that the U.S. Senate resorts to the amendment procedure which is based on a pairwise comparison of decision alternatives with the winning alternative surviving in each comparison to face the next one in accordance with the agenda. When all alternatives have been present in at least one pairwise comparison, the winner of the last comparison is declared the overall winner. With k alternatives one thus conducts $k - 1$ comparisons. This obviously falls dramatically short of the $k \times (k - 1)/2$ comparisons required to directly establish the occurrence or non-occurrence of a cyclic majority. As the voting records only reveal the voting behavior of each senator in various pairwise votes, it is not possible to ascertain all the respective individual preferences relations. Suppose we have three policy options: A, B, and C. If the agenda dictates that the first vote be taken between B and C and the second vote between the winner of this vote and A, we cannot with certainty infer the full ranking of a voter who votes for B in the first and—assuming that it is B that faces A in the second ballot—in the second ballot. We do know that this voter (if voting sincerely) prefers B to both A and C, but we do not know the voter's preference between A and C. Had this voter instead voted for A in the second

ballot, we could infer that the voter's ranking is ABC. Suppose that A represents the *status quo*, i.e. no change, C a legislative proposal and B an amendment to C. Since three alternatives call for two votes, we can denote any voter's voting strategy by an ordered pair (x, y) where $x, y \in \{yes, no\}$ and where the first element indicates whether the voter prefers the amendment to the original motion ('yes' if he/she does, 'no' if he/she doesn't), and the second element indicates whether the voter prefers the winner of the first ballot to the *status quo*. Let $n(x, y)$ be the number of voters with the voting strategy (x, y). Bowen derives four necessary conditions for the emergence of what he calls the voting paradox (cyclic majority relation):

1. the outcome is A (i.e. *status quo*)
2. $n(no, yes) + n(no, no) > n(yes, yes) + n(yes, no)$
3. $n(yes, no) + n(no, no) > n(yes, yes) + n(no, yes)$
4. combining the two last-mentioned conditions yields $n(no, no) > n(yes, yes)$.

The article proceeds to analyze 111 bills subjected to the roll call vote in the Senate. With probabilistic assumptions applied to the preference relations not directly derivable from the voting records, he concludes that only two bills were associated with a high probability of the voting paradox.

The same problem and approach are pursued by Weisberg and Niemi in the same volume (Weisberg and Niemi 1972). Their findings deviate somewhat from those of Bowen either due to the modeling of the voting procedure or due to the assignment of preferences of some voter groups whose preferences are not directly observable. Perhaps more importantly, Bowen classifies as voting paradoxes those cases where the winning alternative is not a majority (i.e. Condorcet) winner, while Weisberg and Niemi (1972, fn 17) look for vote sequences where the winning alternative(s) would have—in the light of the estimated preference rankings—been defeated by at least one other alternative. The difference can be illustrated by the profile of Table 1 (Nurmi, 1999, p. 28). In this 8-voter profile, both A and B beat C (C being the Condorcet loser), but neither beats the other. Thus, there is no Condorcet winner. At the same time, there is no Condorcet cycle. Thus, Bowen would count profiles like Table 1 as instances of the voting paradox, while Weisberg and Niemi would not as neither A nor B are defeated by any other alternative.

The probability of the voting paradox is also the subject of Bjurulf's (1972) article where the main focus is not so much on the specific probability estimates of encountering a cyclic majority but rather on the way voting blocs affect the paradox probability. In other words, Bjurulf studies the effect of introducing specific types of

Table 1 No Condorcet winner and no Condorcet paradox

3 voters	3 voters	1 voter	1 voter
A	B	C	C
B	A	A	B
C	C	B	A

dependence between voters' opinions on cycle probability. His work uses computer simulations to generate electorates with a small number of blocks (parties) in order to find out the effect of relative sizes of the blocks (equal or unequal) and the degree of conflict between the blocks on the probability of majority cycles.

Shepsle (1972) discusses the significance of impreciseness of political positions in the formation of collective preference relations and their acyclicity. He suggests that considering decision alternatives as risky prospects may explain why expected utility maximizing voters with cycle-generating preferences over certain alternatives may avoid majority cycles over the corresponding risky prospects. Thus, making the alternatives more realistic in the sense of allowing for some impreciseness in the voting setting may do away with the cyclic majority in terms of precise positions. Furthermore, the rules of entry of decision alternatives may explain why proposals involving new policy dimensions are often rejected (e.g. by the chairperson) as non-germane. Thus, one source of voting paradox may play a smaller role in practice than in theory.

In the Diss and Merlin volume, Condorcet's paradox is discussed under the topic of Condorcet efficiency of voting rules. Moreover, the paradox is treated as one of several counterintuitive occurrences potentially encountered in the application of voting procedures. The Condorcet efficiency of a voting rule denotes the probability or—as the case may be—the relative frequency of the rule ending up with the Condorcet winner in the Condorcet domain, i.e. in a set of profiles where such a winner exists. So, for example, if procedure P elects the Condorcet winner in eight profiles out of 100 profiles with a Condorcet winner, its Condorcet efficiency is 0.08.

The Diss and Merlin volume begins with Gehrlein and Lepelley's article (2021) Condorcet efficiency of some common (such as the plurality rule, Borda count, plurality elimination, and approval voting) and some fairly uncommon voting rules (such as negative plurality, negative plurality elimination, and Borda elimination) under IC's and IAC's. As stated above, in the former cultures, each voter's preference ranking is drawn randomly and independently from the set of all possible rankings. In other words, for each voter, each preference ranking is equally probable. In IAC, in turn, all distributions of voters over preference rankings are deemed equally probable. Most of what is known about the Condorcet efficiency is based on these authors' earlier works. The new results of this chapter pertain to the effect of potential abstentions of voters to the Condorcet efficiency of various voting rules. This effect turns out to be significant. As in the no-abstention models, the dependency between the voters seems to play a crucial role in determining the outcomes and, in particular, the Condorcet efficiency of voting rules.

The effect of abstentions is also discussed in the chapter by Diss et al. (2021a). More specifically, these authors focus on weighted scoring rules and their Condorcet efficiency. These rules assign scores to various positions in preference rankings and determine the collective rankings in terms of the score sums of alternatives. In addition to abstention, the authors also consider the effect of indifference in preference relations. The main results pertain to the limit values of the Condorcet efficiency when the number of voters is increased. Surprisingly, the Borda count does not turn

out to maximize the Condorcet efficiency, contrary to what has been shown in many earlier studies that do not consider the possibility of voter abstention.

The article of Diss et al. (2021b) addresses a topic that has received relatively scant scholarly attention, viz. the effect that the "closeness" of the election has on the Condorcet efficiency of various voting rules in large electorates.[3] The closeness here refers to the rule-independent opinion distributions preceding the actual voting. Intuitively, one would expect that the presence of one formidable candidate would increase the Condorcet efficiency of rules that elect strong Condorcet (a.k.a. absolute) winners, such as plurality, plurality runoff, instant runoff or Bucklin's rules. Hence, in close elections, a wild guess would be that the effect is the opposite: the closer, the less Condorcet efficient. Using a specific index of election closeness, the authors show that this wild guess is, indeed, a guess: it holds for some rules, but not for all. A notable exception is the Borda count under which the (high) Condorcet efficiency seems to remain largely unaffected by the closeness. It should be added, though, that the results hold for three-candidate elections only.

Brandt et al. (2021) deal with two paradoxes: the Condorcet loser paradox (the eventual election of a Condorcet loser) and the agenda contraction paradox. The former occurs when a preference profile is found or (more commonly) constructed so that the voting rule under investigation results in an alternative that would be majority defeated in pairwise contests by every other alternative. The latter paradox occurs when it turns out that, in a given profile, the rule under study would result in a given alternative, but if some alternatives other than the given one were removed from the profile, another alternative would emerge as the winner.[4] Brandt et al. introduce a (in voting contexts) new methodology, Ehrhart theory, for the analysis of these paradoxes in four-alternative situations mainly under the IAC assumption. The article concludes that the probability of encountering an instance of the Condorcet loser paradox when resorting to some well-known Condorcet extensions (MaxiMin, Dodgson, Tideman, Young) is so low as to make it practically irrelevant under the cultures studied. So, in the authors' view the vulnerability to the Condorcet loser paradox cannot be used as grounds for dismissing those extensions. On the other hand, the occurrence of the agenda truncation paradox seems far more likely.

The picture emerging from the above chapters in Diss and Merlin (2021) is much more nuanced than the one outlined in the chapters of Niemi and Weisberg (1972) touched upon above. At the same time, Diss and Merlin (2021) provides a wider setting for the study of paradox of cyclic majorities. Indeed, Gehrlein and Merlin (2021) analyze the probability of Ostrogorski's paradox, while Belayadi and Mbih (2021) focus on the probability of reversal symmetry violations under various voting rules. The former paradox is related to aggregation of entries in a $k \times m$ matrix of a's and f's each entry indicating, say, that the candidate or applicant for a job qualifies (does not qualify, respectively) on criterion i in evaluator j's opinion with $i = 1, \ldots, k$ and $j = 1, \ldots, m$. In this case, the entry $(i, j) = a$ ($(i, j) = f$, respectively). Sup-

[3] This approach differs from (in fact, reverses) the one where the effects of voting rules on "closeness" of results are sometimes analyzed.

[4] This paradox is also known as the violation of the Chernoff or heritage property.

Table 2 Strong Ostrogorski's paradox

	Criterion 1	Criterion 2	Criterion 3	Aggr.value
Evaluator 1	f	f	a	f
Evaluator 2	f	a	f	f
Evaluator 3	a	f	f	f
Evaluator 4	a	a	a	a
Evaluator 5	a	a	a	a
Aggr.value	a	a	a	a or f

posing that each evaluator considers each criterion equally important, it makes sense to summarize each evaluator's opinions as either a if his/her evaluations of the candidate have more a- values than f-values and vice versa. Considering all evaluators we thus have m summary entries each either a or f. Using the same principle, we can now derive an aggregate evaluation of a candidate by assigning the candidate the value a (f, respectively) if there are more a's(f's) in the m columns. This would amount to columns-first aggregation.

Similarly, one could resort to the rows-first aggregation and derive the aggregate value for the k criteria. Now it may happen that one ends up with contradictory overall evaluation so that the columns-first aggregation leads to value a and the rows-first to f or vice versa. Should this happen, we have an instance of Ostrogorski's paradox. If one of the conflicting evaluations results from 'unanimous' aggregate evaluations (either all evaluators give the same aggregate assessment of the candidate or all criteria suggest the same aggregate value for the candidate), then we have an instance of the strong Ostrogorski's paradox. Table 2 where $k = 5$ and $m = 3$ illustrates.

Here a job applicant is being evaluated by five evaluators, each assigning him either 'pass' (denoted by 'a') of 'fail' (denoted by 'f') on each criterion. The overall mark of the applicant given by the evaluator is obtained as the value that most often occurs in this evaluator's criterion-wise evaluations of the candidate. Similarly, the criterion-wise mark of the applicant is obtained as the value given by the majority of the evaluators on this criterion. These values are presented in the right-most column and lower-most row in Table 2. The paradox consists in the observation that the two ways of aggregating values lead to conflicting results: rows-first yields 'f', while columns-first gives 'a'. Note that the latter conclusion stems from, not just from a majority, but from all aggregated criterion-wise values suggesting it. In fact, 60% of the voters vote for 'a' on each criterion and, yet, 60% of voters support 'f' on the basis of their aggregated evaluations. Gehrlein and Merlin find that the strong version of Ostrogorski's paradox is extremely unlikely to happen under the IC assumption.

Reversal symmetry paradox (a.k.a. preference inversion paradox or reversal bias) occurs when, under a given voting rule, the reversal of a preference profile leads to the same outcome as the outcome ensuing from the application of the rule in the original profile. Table 3 gives an instance of the paradox under the plurality runoff

Table 3 Reversal symmetry paradox and plurality runoff

4 voters	3 voters	2 voters
A	B	C
B	C	A
C	A	B

system.[5] In the initial profile, there is a runoff between A and B, whereupon A wins. In the profile where all rankings are inverted, we get a runoff between A and C where A again comes out victorious. Hence, the same outcome ensues in a profile and its complete inverse.

Belayadi et al. derive closed formulae for the probability of reversal symmetry paradox in IC's as well as in IAC's for selected voting rules for a small number of alternatives (3 and 4). It seems that the results largely support the contention of Regenwetter et al. (2006) according to which IC tends to exaggerate paradox probabilities. It should be born in mind, though, that while the IAC seems in general to be associated with lower paradox probabilities than the IC, the finding doesn't say much about the likelihood of encountering the said paradoxes in real world voting situations. We shall return to this point later on.

3.2 The U.S. Political Institutions

In the 1970s, the probabilistic modeling was largely conducted in the U.S. universities. This is reflected in the author list of Niemi and Weisberg (1972): only two authors out of 17 had a non-American affiliation. In the Niemi and Weisberg volume two U.S. institutions—the Congress and the Supreme Court—are singled out. Bowen's contribution was already discussed above. Its main focus is in the possibility of cyclic majorities underlying the votes in the Senate. Koehler (1972) examines the coalition formation in the U.S. House of Representatives using as the benchmark the size principle, i.e. the hypothesis that coalitions of minimal winning size will form. In the words of Riker (1962, pp. 32–33):

> In social situations similar to *n*-person, zero-sum games with side-payments, participants create coalitions just as large as they believe will ensure winning and no larger.

Using the Brams-O'Leary index of voting agreement (Brams and O'Leary 1970), Koehler examines the roll call votes in eleven sessions of the House of Representatives to assess the predictive value of the size principle. Both party and individual level agreement are calculated. The concept of probability enters the discussion via the Brams-O'Leary index which is based on expected numbers of voting agreement. Rohde's focus is on the validity of the size principle, but in a context where zero-sum assumption underlying Riker's construction seems at first blush somewhat bizarre, viz. in the U.S. Supreme Court decision-making (Rohde 1972). What the size principle would then imply in the nine-member Court is that a five-member opinion

[5] The profile is a minor simplification of the one presented in (Felsenthal and Nurmi, 2018, p. 34).

Table 4 An instance of the referendum paradox (Nurmi, 1999, p. 77)

Ref. stand	MP1	…	MP167	MP168	…	MP200	Row sum
Yes	7.000	…	7.000	15.000	…	15.000	1.664.000
No	8.000	…	8.000	0	…	0	1.336.000

coalitions be formed. However, Rohde argues that in decisions under an external threat to the Court, the coalitions will be larger than minimal winning. Assuming that each judge votes 'yes' or 'no' with probability 1/2, Rohde computes the theoretical probabilities that majority coalitions of various sizes form. He then compares empirically observed coalition data with these values. He finds that there is a significant difference in the probabilities predicted by the size principle and the empirical data. So, the principle doesn't seem applicable to the Supreme Court decision-making either in the case of opinion situations or under external threat. He concludes that a more nuanced study of various decisions situations is called for in order to find out the domains where the size principle is applicable as a predictor.

Now, some 50 years later, the probabilistic study of voting rules has a different focus when it comes to political institutions. Instead of the study of roll call votes in the main U.S. federal institutions, the idea of federalism itself is brought to the agenda of study in the Diss and Merlin volume. Specifically, two multi-author contributions focus on the representative principles in federations. Feix et al. (2021) study two-tier systems of governance where on one tier the voters elect representatives for an assembly that constitutes the second tier. The ultimate decision maker is the latter tier. Numerous institutions are instances of the two-tier arrangement: the Council of Minister of the European Union and the Electoral College of the U.S. being perhaps the best-known ones, but basically all representative institutions are two-tier systems: the voters elect parliamentarians (or other representatives) to make binding decisions on legislation, policies, allocation of resources, etc. Because the choices (decisions) made on the second (representative) tier are not ones voted upon on the first tier and yet the voters on the latter often have opinions regarding those choices as well, it may happen that the choices ensuing from the decision making of the representatives would be voted down by the electorate at large had it been given the opportunity to vote directly on those choices. This kind of occurrence is sometimes called the referendum paradox. A fictitious example from Nurmi (1999) illustrates. A parliament consists of 200 members (MP's) each elected from a single-member constituency. The voting population is assumed to be 3 million. This is distributed into 200 constituencies of equal size, viz. 15000 voters. The country ponders upon entering an economic union. The stands of the voters in each constituency are presented in Table 4. Should a referendum be arranged, 'yes' would clearly be the victorious stand with more than 55% of the voters voting for it. However, should the issue be decided in the parliament, a whopping 5/6 of the MP's would 'in good faith' support 'no' correctly convinced that they are expressing the will of the majority of their constituency.

Feix et al. evaluate various voting rules in terms of a criterion they have devised, viz. majority efficiency. A method is the more majority efficient the less likely it makes the instances of referendum paradox to arise. The article discusses the Penrose–Banzhaf (independent population) probability model (Penrose 1946; Straffin 1977) and compares it with May's (homogeneous population) model (May 1948; Straffin 1977) to find out whether the allocation of seats to states in accordance with Penrose's proposal, i.e. in proportion to the square-roots of the respective populations, leads to a more majority efficient allocation than May's model that suggests seat allocation in direct proportion to the populations. The authors resort mainly to Monte Carlo simulations to estimate the majority efficiencies of various allocation rules. They also study rules that assign seats to states according to the formula

$$a_i = n_i^{\delta}$$

where a_i is the number of seats allocated to state i, n_i the population of state i and δ a coefficient ranging from 0 (one seat per state) via 1/2 (the square-root rule) and 1 (strict proportionality) to ∞ (all seats to the most populous state). For each situation that includes a fixed number of states and decision alternatives, the authors run 10^5 or 10^6 simulated elections under IC and IAC assumptions to find out the majority efficiencies of various rules. It turns out, e.g. that the square-root rule is less obvious winner in IC cultures than has been thought.

The article of De Mouzon et al. (2021) can be considered a companion article of Feix et al. (2021) in the sense that both of these articles make the same probabilistic assumptions regarding the electorate. The central concept studied is the justice of the election procedure in two-tier systems. This contribution utilizes the standard power indices—the Banzhaf and Shapley–Shubik indices—and looks at the distributions of these values within each state in order to judge the justice of the system. A whopping 10^{12} simulation rounds were conducted in each state of the U.S. to obtain estimates of pivot probabilities of voters in each state. Owen's study from 1970's is used as a benchmark in comparing deviations from perfectly equal pivot probabilities between the states Owen (1975). It turns out that the main conclusions regarding electoral justice hinge on the probabilistic population models. Specifically, May's model leads to the conclusion that violations of electoral justice mainly take place at the expense of populous states, while both Banzhaf and Shapley–Shubik probability models suggest the opposite, i.e. that the deviations occur mainly to the disadvantage of small states.

4 Non-overlapping Themes

The preceding section singles out a couple of topics discussed in the two volumes reviewed here. Apart from these, there are several themes that are discussed in one of the books only. One of these is coalition formation and another spatial political competition. This doesn't mean that study of these areas has stopped over the 50 years that followed the publication of the Niemi and Weisberg volume. Coalition formation

remains an important topic in basically all multi-party democracies. Moreover, Diss and Merlin (2021) has, as its title suggest, a probabilistic emphasis, while most political coalition theories are deterministic. Nonetheless, one cannot escape the impression that the probabilistic study of voting rules is today mainly in the domain of other branches of science than political science.[6]

The Diss and Merlin book contains a host of topics not touched upon in the Niemi and Weisberg one. Given the time lapse between the two books, this is only natural. Much of the work done on voting rules is based on problems left open by earlier studies. In fact, this is the way many new voting rules are invented. However, many open problems can be addressed with methods borrowed from disciplines not directly related to voting. Probability modeling is, of course, a good example. However, in addition to probability theory and statistics, computer science has entered the field in a big way creating a new sub-field in voting theory: computational social choice. This development has its origins much later than the Niemi and Weisberg volume. As probability modeling, the computational social choice has to confront the question of relevance: what can be achieved through the use of the tools and techniques stemming from computer science? In the case of probabilistic modeling the answer was to give a more nuanced picture of the virtues and vices of voting rules. Instead of just[7] stating that a procedure fails on a criterion of performance, we need to have some idea about how likely such failures are. Probability models provide—among other things—relevant information insofar as they give estimates about the failure probabilities and—perhaps more importantly—about factors increasing or decreasing the failure probabilities. Similarly, the computational techniques provide information about the practical difficulty of finding the winners or about the difficulty of successfully misrepresenting one's preferences and similar aspects affecting the relevance of the voting theory results. In a similar vein, Brandt et al. (2021) apply the computational approach of Ehrhart (1962) to find out the likelihood of the no-show paradox. Also, Karpov (2021) provides an overview of combinatorial methods in analyzing voting rules. Moyouwou et al. (2021) present an overview of the pioneering Fishburn–Gehrlein method for computing the number of integer solutions to a set of linear inequalities, a method that has subsequently been often applied, notably by Gehrlein and his associates, in probabilistic voting studies. The authors also discuss Ehrhart's conjecture and provide outlines for associated algorithms. Overall, the menu of techniques is thus considerably more extensive than in the Niemi and Weisberg book.

When it comes to topics not present in the earlier volume, two parts of Diss and Merlin (2021) stand out: one on resistance to manipulations and the other on game theory. An obvious reason for the absence of the former theme in the Niemi and

[6] This is not the proper place to address the age-old topic of what constitutes a science *sensu stricto*. No doubt voices have occasionally been raised to argue that 'science' should denote natural sciences and should not be applied to the study of such 'soft' entities as politics—or economics for that matter.

[7] This expression is not intended to play down the significance of findings regarding paradoxes or incompatibilities between performance criteria. Such findings may be exceedingly difficult to make.

Weisberg volume is that the idea of strategic voting and the associated outcome manipulability had not been widely studied at the time of the publication of the volume.[8] The part on game theory focuses on rather specific topics in game theory rather than on the general foundations of voting games. The Diss and Merlin volume concludes with a progress report on probability calculations based on the IAC assumption. This assumption was not used in the Niemi and Weisberg volume where the dominant assumption is IC.

5 Progressive Program

All this leads to the conclusion that the probability modeling of voting rules is what used to be called a research program in the sense of Imre Lakatos: a theoretical-empirical system with a distinctive set of core assumptions not challenged within the program and surrounded by lower-level theoretical results challenged occasionally with empirical evidence Lakatos (1977). Should serious discrepancies between the obtained evidence and theoretical statements or principles ensue, those lower-level statements are dispensed with or replaced with others, while the core tenets are maintained. The probability modeling tradition fits these hallmarks of a research program. It is also a progressive entity: in time, the theoretical foundation gets more nuanced and the range of procedures and their workings expand so that the new structures embrace the older ones in accounting for the successes of the latter while correcting their failures. In the probability tradition, the core beliefs pertain to the basic constituents: the culture assumptions, the individual preferences, the voting strategies, and the outcomes.

The basic finding in the comparison of the two volumes is that the models and research strategies have become far more nuanced and sophisticated with new insights from other disciplines, mathematics and computer science, in particular. Progress is also visible in the developments of computational capacity of modern machinery. What in the 1970s seemed like an adequate number of (simulated) elections to guarantee reasonable convergence of the estimated probabilities, say about 10^4 rounds of elections, required a considerable amount of computational resources limiting the simulation modeling to relatively small candidate and voter sets. Nowadays, these restrictions have been relaxed to a substantial degree. Importantly, also the algorithms used in solving various problems through simulations have become more versatile. In short, the probability modeling has made progress in theoretical, methodological, and technical aspects. At the same time, the small community of scholars contributing to the Niemi and Weisberg volume has grown to a vast network of researchers from all continents.

[8] See, however, the pioneering work of Farquharson (1956) where the idea of sophisticated voting as a form of successful preference misrepresentation and the concept of straightforward voting rule as a rule immune to successful misrepresentation was introduced. These ideas were further developed in Farquharson (1969). See also Dummett and Farquharson (1961).

6 What Do We Learn from Probability Models?

In what way do the probability models of voting rules then help us in understanding voting institutions? A partial answer to this is that these models stand midway between the compatibility/incompatibility results of the social choice theory and empirical data on voting results. This is nicely expressed by the two honorees of the Diss and Merlin volume in the quote of Sect. 2 above. The probability models transform basically dichotomous evaluations into graded ones. To put it bluntly, they tell us which ones of bad procedures are less bad than others in terms of specific criteria. This sounds somewhat akin to Tolstoy's famous dictum:

> Happy families are all alike; every unhappy family is unhappy in its own way.

Translating this to the language of voting rules, the happy ones would all satisfy a criterion of goodness. For a voting rule being unhappy would then mean that it fails on a theoretical criterion and being unhappy in its own way would mean that it has its own probability of failing. So, the picture emerging from probability estimates is more precise than the one resulting from studies on the compatibility of norms. From a practical point of view, various 'cultures', e.g. IAC, are profile restrictions and, as Arrow famously showed, restrictions make a great deal of difference in the compatibilities of norms. Accordingly, to rely on estimates based on IC amounts to using irrelevant information in institution design in situations where it is known that some fixed restrictions always hold. Thus, for example, the domain of single-peaked preferences or, more generally, the Condorcet domain is of particular interest in cases where the preferences tend to obey this restriction. The difficulty is that the set of potentially relevant sub-domains is not known beforehand.

The upshot is that the probability estimates of paradoxes do not say much about the likelihood of encountering those paradoxes in practice. After all, we are dealing with statistical inference with at best confidence degrees and intervals. So, the judgments regarding voting rules are bound to be of the type: 'rule A leads to a smaller failure probability on criterion X than rule B assuming that the relevant voter set is drawn randomly from a population of profiles satisfying assumption Y'. This leaves open the possibility that the priority of A *vis-à-vis* B is inverted in some proper subset, say Y', of those profiles satisfying Y. This in itself is natural and even obvious, but in cases where Y' is the profile restriction prevailing in practice, we may be misled in our designs based on probabilities based on Y.

7 By Way of Conclusion

The two volumes reviewed above open a rich panorama of the developments in the field of probability modeling of voting rules. Over the 50-year span, the progress has been dramatic in terms of methodologies employed and in terms of the subjects included in the study. Yet, some crucial features of the Niemi and Weisberg volume

remain visible in the Diss and Merlin book. What started as a study of the probability of ending up with cyclic majorities in specific voting bodies has over time developed into a technically sophisticated analysis of many other types of paradoxes. The likelihood of encountering profiles with a Condorcet winner has been complemented with the study of Condorcet efficiency of non-Condorcet voting rules. The assumption of sincere voting has been replaced by a more nuanced view of voter motivations. Accordingly, manipulation probabilities have become the focus of a distinct body of literature. The range of institutions studied has widened from committee-type bodies to representative two-tier systems. These are but some of the most prominent developmental strides so far taken on the way to a better understanding of voting rules from the probabilistic perspective.

References

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). Wiley.

Belayadi, R. & Mbih, B. (2021). Violations of reversal symmetry under simple and runoff scoring rules, pp. 137–160 in Diss & Merlin (2021).

Bjurulf, B. (1972). A probabilistic analysis of voting blocks, pp. 232–251 in Niemi & Weisberg (1972).

Black, D. (1958). *Theory of committees and elections*. Cambridge University Press.

Brams, S. J., & O'Leary, M. (1970). An axiomatic model of voting bodies. *American Political Science Review, 65*, 449–470.

Bowen, B. D. (1972). Toward an estimate of the frequency of occurrence of the paradox of voting in U.S. Senate roll call votes, pp. 181–203 in Niemi & Weisberg (1972).

Brandt, F., Geist, C. & Strobel, M. (2021). Analyzing the practical relevance of the Condorcet loser paradox and the agenda contraction paradox, pp. 97–115 in Diss & Merlin (2021).

Brandt, F., Hofbauer, J. & Strobel, M (2021). Exploring the no-show paradox for Condorcet extensions, pp. 251–273 in Diss & Merlin (2021).

De Condorcet, M. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. L'Imprimerie Royale.

De Mouzon, O., Laurent, T., Le Breton, M. & Moyouwou, I. (2021). "One man, one vote" part 1: Electoral justice in the U.S. Electoral College: Banzhaf and Shapley/Shubik versus May, pp. 189–227 in Diss & Merlin (2021).

Diss, M., Kamwa, E., Moyouwou, I. & Smaoui, H. (2021a) Condorcet efficiency of general weighted scoring rules under IAC: Indifference and abstention, pp. 55–73 in Diss & Merlin (2021).

Diss, M. & Merlin, V. (Eds.). (2021). *Evaluating voting systems with probability models*. Essays by and in Honor of William Gehrlein and Dominique Lepelley. Springer.

Diss, M., Pérez-Asurmendi, P. & Tlidi, A. (2021b). The effect of closeness on the election of a pairwise majority rule winner, pp. 75–95 in Diss & Merlin (2021).

Dummett, M., & Farquharson, R. (1961). Stability of voting. *Econometrica, 29*, 33–43.

Ehrhart, E. (1962). Sur les polyèdres rationnels homothétiques à *n* dimensions. *Compte-Rendus de l'Academie des Sciences, 254*, 1273–1302.

Farquharson, R. (1956). Straighforwardness in voting procedures. *Oxford Economic Papers, 8*, 80–89.

Farquharson, R. (1969). *Theory of voting*. Yale University Press.

Felsenthal, D. S., & Nurmi, H. (2018). *Voting procedures for electing a single candidate*. Springer.

Feix, M., Lepelley, D., Merlin, V., Rouet, J. - L. & Vidu, L. (2021). Majority efficient representation of the citizens in a federal union, pp. 163–187 in Diss & Merlin (2021).

Gehrlein, W. V., & Lepelley, D. (2004). Probability calculations in voting theory: An overview of recent results. In M. Wiberg (Ed.), *Reasoned choices* (pp. 140–160). The Finnish Political Science Association: Helsinki.
Gehrlein, W.V. & Lepelley, D. (2021). Analyzing the probability of election outcomes with abstentions, pp. 15–53 in Diss & Merlin (2021).
Gehrlein, W. V. & Merlin, V. (2021). On the probability of the Ostrogorski paradox, pp. 119–135 in Diss & Merlin (2021).
Gibbard, A. (1973). Manipulation of voting schemes. *Econometrica, 41*, 587–601.
Holler, M. J. (1978). A priori party power and government formation: Esimerkkinä Suomi. *Munich Social Science Review 1*, 25–41.
Holler, M. J. (Ed.). (1982). *Power, voting and voting power*. Physica-Verlag.
Holler, M. J. & Rupp, F. (2020). Some strategic decision problems in networks. *Transactions of Computational Collective Intelligence XXXV*, 133–147.
Karpov, A. (2021) Combinatorics of election scores, pp. 347–366 in Diss & Merlin (2021).
Koehler, D. H. (1972). The legislative process and the minimal winning coalition, pp. 149–164 in Niemi & Weisberg (1972).
Lakatos, I. (1977). *The methodology of scientific research programmes: Philosophical papers* (Vol. 1). Cambridge University Press.
May, K. O. (1948). Probabilities of certain election results. *The American Mathematical Monthly, 55*, 203–209.
Moulin, H. (1988). Condorcet's principle implies the no-show paradox. *Journal of Economic Theory, 45*, 53–64.
Moyouwou, I., Andjiga, N. G. & Mbih, B. (2021). From Gehrlein-Fishburn's method on frequency representation to a direct proof of Ehrhart's extended conjecture, pp. 367–398 in Diss & Merlin (2021).
Nanson, E. J. (1883). Methods of election. *Transactions and Proceedings of the Royal Society of Victoria*, art. XIX, 197–240.
Niemi, R. G., & Weisberg, H. F. (Eds.). (1972). *Probability models of collective decision making*. Charles E. Merrill.
Nurmi, H. (1999). *Voting paradoxes and how to deal with them*. Springer Verlag.
Owen, G. (1975). The presidential election game. *American Political Science Review, 69*, 947–953.
Penrose, L. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society, 109*, 53–57.
Regenwetter, M., Grofman, B., Marley, A. A. J., & Tsetlin, I. (2006). *Behavioral social choice: Probabilistic models, statistical inference and applications*. Cambridge University Press.
Riker, W. H. (1962). *The theory of political coalitions*. Yale University Press.
Rohde, D. W. (1972). A theory of the formation of opinion coalitions in the U.S. Supreme Court, pp. 165–178 in Niemi & Weisberg (1972).
Satterthwaite, M. (1975). Strategy-proofness and Arrow's conditions. *Journal of Economic Theory, 10*, 187–217.
Sen, A. K. (1970). The impossibility of a Paretian liberal. *Journal of Political Economy, 78*, 152–157.
Shepsle, K. A. (1972). The paradox of voting and uncertainty, pp. 252–270 in Niemi & Weisberg (1972).
Straffin, P. D. (1977). Homogeneity, independence and power indices. *Public Choice, 30*, 107–118.
Todhunter, I. (1865/1949). *The history of the mathematical theory of probability from the time of Pascal to that of Laplace*. Chelsea Publishing Company.
Weisberg, H. F. & Niemi, R. G. (1972). Probability calculations for cyclical majorities in Congressional voting, pp. 204–233 in Niemi & Weisberg (1972).

The Liblice Castle Story

Dušan Tříska

1 Introductory Notes

1.1 Objectives

Section 2 of this essay can be taken as a case study on the relationship between political *power* and *responsibility*. It is based upon the author's experience from the 1990s—his participation in designing and implementing the post-communist institutional change of Czechoslovakia and, somewhat later, Czech Republic.

In Sects. 3 and 4, the author changes the genre and turns to a *positive* (value-free) societal scholarship or, put differently, the science for its own sake.[1] Extending upon his Tříska (2017), he will seek to remind scholars of their task to find agreement on the *universal analytical building blocks* of social science.[2] Here he dares to refer to his own proposal, namely, the blocks by which he has attempted to bridge legal scholarship and economics.

However, the genuine contribution to the Festschrift can be found in its extensive footnotes through which he sought to associate his ideas to the unique relationship with Manfred J. Holler (further only "MJH").

[1] We will use the term *positive* in the sense of Friedman (1953).

[2] The call for *universal analytical building blocks* is due to Ostrom (2006, 2009).

D. Tříska (✉)
Prague University of Economics, CEVRO Institute, Nove Mesto, Czechia
e-mail: nf@vse.cz

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_12

1.2 A Historiographical Note

To begin with, we must go back to March 1989 when at Liblice Castle—some 25km north of Prague—a few local enthusiasts managed to organize an international (East–West) economic conference *Micro-models*.[3]

In the time passing the event became a legend, albeit primarily for the political import of the fact that—after decades—the Conference was allowed to ever happen. Notably, already this—as if—trivial breakthrough could have been—already then—read so that the days of the Soviet empire had been numbered and the Central and Eastern Europe (CEE) will be soon mercilessly thrown into institutional chaos euphemistically called *transition*.

Yet, from the perspective of the organizer's self-interests, the objective of the Conference was not to pull down Berlin Walls or initiate the Velvet Revolution. What they were after was only to constitute of a communication platform with western scholars, such as happened to be the Conference participant named MJH.

As indirectly suggested, the new East–West platform did not survive more than a few months as many of the CEE participants got lost in the political turmoil. To illustrate, this essay's author soon became a deputy of the Federal Minister of Finance Václav Klaus ("VK")—another participant of the Conference, what else.[4]

1.3 A Methodological Note

For the sake of this essay, we will accept that the dividing line between *economics* and *political economy* is defined by their descriptive (positive, value-free) and normative nature, respectively.

To illustrate, in March 1989 it would be unthinkable to invite papers with a normative content to Liblice. Hence, it was "only the ordinary" microeconomics that was practiced at the Conference – often in the format of "mathematical" comparative models of planned and market systems.

Extending upon the author's major monograph,[5] Sects. 3 and 4 amount to a few elements of the value-free analysis. Until then, right in the next Sect. 2, the normative notion of *responsibility* will be exemplified by the value-loaded work for VK's Finance Ministry.[6]

---

[3] Cf. proceedings of the Conference Quandt and Tříska editors (1990).

[4] For completeness, let us recall that VK later became prime minister of the Czech Republic and in the years 2003–2013 served as the country's president. Hence, it must have been an exceptional foresight that made MJH bring VK—right after the Conference—to Aarhus University. In the same mode, MJH intermediated the publication of Tříska (1989), notably the author's first ever publication in the West.

[5] Tříska (2017).

[6] Section 2 extends upon the author's course *Economics and Politics of Institutional Change* at the CEVRO Institute in Prague. The nature of his lectures can be well seen in this essay's bibliography.

2 Political Economy of an Institutional Change

2.1 Policy/Responsibility No. 1: Transformation Versus Reform

In the Finance Ministry, we became primarily responsible for the verdict that the economy based on commands is patently inferior to that coordinated by the Invisible Hand of the Market–put in the very brief.

Consequently, then, we also had to publicly admit that it would not be enough to only *reform* the economy as the economic system must be fundamentally *transformed*. In other words, we simply had to publicly declare that the core of our plan lies in the Schumpeterian *creative destruction* of the existing institutional setting - to open the way for the new and *supposedly* superior social-economic order.

The notion of *responsibility* had a special meaning to us due to our—by and large—libertarian background, namely, the imperative that whenever a politician may dare evaluate an institutional setting as being "better", he-she should immediately qualify who may end up relatively worse-off under this "superior" setting.

2.2 Policy/Responsibility No. 2: Social Order "By Design"

We were fully aware of Hayek's observation that: "…the curious task of economics is to demonstrate to men how little they know about what they imagine they can design…[7]

In the perfect opposite to this, we had to respond the public demand for detailed blueprints of our strategy, preferably in the format of a workflow chart representing exact *timing* and *sequencing* of our institutional measures, e.g., of the order in which we will liberalize prices and adopt a flexible exchange rate.

Similarly attached we were then to Frank Knight's school of thought—namely, his ethical skepticism[8] and recommendations what the economists should or should not do.[9]

---

[7] Cf. Hayek (1991).

[8] Cf. Knight (1982a, 1982b, p.7): "… the very notion of freedom as a "criterion" is illusory. The theory of maximum freedom … ends in a question-begging justification (fallacy of petitio principii) …, and the only way to escape this result and arrive at any ethical judgment is to appeal to an ethical judgment as to such. ….. If the notion of freedom has an ethical significance, it is derived from prior ethical norms".

[9] It must have been in 1999 that MJH endowed me with a hard-cover copy of Buchanan (1964)—extensively quoted in this essay. This event is to demonstrate here that 10 years after the Liblice Conference our relationship remained on the level of exchanging presents. Honestly speaking, during my hectic times in politics, the Hollers (Manfred, his parents, and his wife Barbara, also an economist) provided me—and my wife Hana, a sinologist—with a hideout at the Gnesener Str. 1, Daglfing, Munchen. In 2000 James Buchanan visited Prague and I made him increase the book's market value by his dedication to MJH, me, and our wives.

Wasn't it then—in principle—somewhat *irresponsible* from classical liberals to ever accept the task to "run their country"?[10]

2.3 Policy/Responsibility No. 3: Variant Targets of the Institutional Change

Trivially, a transition process is about transferring the country from "somewhere" to "somewhere else". The transition's starting point was—after August 1968—labeled as a *real* socialism to strictly differentiate the system from the reformists' dreams about socialism "with a human face".

As to the transition's target, our firm position was that it must belong to the category of a *democratic capitalism*. Even today, lively debates are sure to arise about the system's specific attributes of this thus defined target, about the similarities and differences between the German-style *Soziale Marktwirtschaft* and, e.g., the capitalism practiced in Japan, Singapore, and South Korea.

Still more intriguing could then be the fact that the pattern selected as our *target* will certainly be *moving*. Hence, we could only speculate into what form will mutate our Thatcher-Reagan ideal by the time we will have completed the transitory mission.[11]

However uncertain may have been our ideas about democratic capitalism, we concluded that: firstly, a maximum support must be provided to the emerging private entrepreneurship and, secondly, the fatal obstacle here rests in the *institution* of a state-owned enterprise (SOE). As a result, the inherited SOEs had to become the No. 1 "victims" of the *creative destruction*. The name of this destruction is *privatization*—to be discussed later in Sect. 3.

2.4 Policy/Responsibility No. 4: Efficiency Versus Justice/Fairness

As a rule, political choices bring forward different costs and benefits to different social groups. In particular, different privatization strategies are likely to turn different categories of citizens into ultimate *winners* and *losers* of the institutional change. The new reallocation of wealth and power may thus lead to a political explosion if

[10] It was MJH's Barbara who used the term run to characterize my attempts to restructure the CEE.

[11] To illustrate, the laissez-fair market of the nineteenth century metamorphosed into what Holcombe (2015) is ready to call *political capitalism*. While a libertarian of my caliber certainly does not welcome this development. I must admit that I can only speculate what are MJH's feelings about, e.g., today's globalization. However, the hard data is that MJH asked me to write for his Homo Oeconomicus the comment Tříska (2016) on Phelps (2015), where this Keynesian Nobel Prize laureate—to my amazement—glorifies innovativeness of the 19the century economy and launches almost Knightian attack on the western intellectuals' elitism.

regarded as unfair, or even a consequence of criminal activities of the transformation's architects.

Leaving aside the criminal aspect of the problem, what comes to fore is the methodological puzzle of the trade-off between the transition's economic and *non-economic* consequences, i.e., between economic efficiency and, on the other hand, e.g., *equality*, *fairness*, *liberty*, and *justice*.[12]

This said, it is noteworthy that the non-economic dimension of the trade-off is—as a rule—by far less understood than the economic one. Essentially, it is beyond human knowledge to *ex ante* rank policies according to how *fair* or *unfair* they will be seen in the short and long run. Whereas economic valuations can be—to some extent—based on relatively "rational" arguments, the non-economic value judgments are often products of "mere" emotions.[13]

Hence, it should be of no surprise that the harshest political disputes—at home and internationally—have concerned the stubborn disparity between the *revolutionary* speed with which market forces embrace the society and the immeasurably longer time needed by the *evolution* of the legal frame, its design and enforcement.

As already suggested, at stake are here requirements upon timing and sequencing of the respective institutional measures, in an implementation of an "adequate" regulation of the business-type activities of the new-born capitalists. As to this, the prevailing thesis then and still today has stated that the privatization should be postponed until the "rule of law" was in place.

Our counter-argument then was—and still is—that it is unrealistic to expect a post-communist government to mobilize *power* strong enough to *postpone* the process whose roots must be traced to the historically unique and unexpected geopolitical explosion—the collapse of the Soviet empire.

It may be of value to recall that the CEE countries were to impose their *creative destruction* not only on the economy but also on the government itself, including the necessity to somehow resolve the problem with the exponents of the former system.[14] Closely associate with this task is the fact that the government desperately searched for experts who would be able and willing to implement *into* the existing system the completely new agendas, e.g., those on bankruptcies, collective investment, insider trading, minority share-holders protection, mergers, and acquisition.[15]

---

[12] The readings for my lectures at CEVRO include concepts of a trade-off discussed, e.g., in Okun (1975), Buchanan (1993), Kaplow and Shavell (2009).

[13] Cf. Knight (1982a, 1982b, p. 155): "Economic thought runs almost entirely in terms of the obvious and commonplace, while political thought is almost as exclusively inchoate, indefinite, and inconclusive, and in consequence political opinion is a matter of wish-thinking and romanticism in overwhelming variety."

[14] The issue of the so-called transitory justice has been left aside in this essay and so were our debates at the Holler's over the similarities and differences between decommunization and denazification.

[15] Right after the visit of Margaret Thatcher to VK's Finance Ministry (June 1990), an advisory team was selected in London—with privatization as its major agenda. It soon became clear that there is a fundamental difference between the objectives the western and eastern version of—as if the same—process. Alongside this team sponsored by the accountants, bankers, and lawyers sponsored by the British Know-How Fund, my private know-how base had its headquarters in the Bavarian

And still, even today many observers can see it as patently *irresponsible* from us to open markets in what then could resemble legal vacuum.[16] Their advice was—let us recall—to interrupt the avalanche to "buy time" until the proper institutions could be imported, safely installed, and reliably tested for their efficiency.[17]

Let it be then stressed again that the government was not strong enough to block the market's opening even if it wanted to do so.

2.5 Policy/Responsibility No. 5: Scholar Turned Politician

On the face of it—as already stated—it appears unlikely that a "true" scientist would ever agree to exchange his-her academic career for politics. Hence, we should ask what may have been then the attractions of politics to the men and women of science, including the organizers of the Liblice Conference?

Leaving aside the Mother Tereza kind of *responsibility* for public good, the first candidate for an honest answer could be the *curiosity* as the fundamental characteristic of a true scientist. Some of us simply wanted to see their subject matter *in action.*

However, our answer will be different. We will claim that the strongest attraction of political power for scientists, including the author of this essay, rests in their desire to see their *intellectual production come true*—realized in practice. Needless to stress that the desire increases with the historical uniqueness of the circumstances and that the social-economic earthquake of the 1990s was an extremely unique opportunity—to be offered again in decades if not centuries.

What we try to invoke here is the well-known libertarian maxim about intellectuals (scientists, technocrats, experts of all kinds) to be a highly dangerous breed of politicians, who—if in power—may want to see realized—cost what it may—their *inventions*—their social engineering plans for a better and fairer society.

Consequently, it is primarily this breed of politicians—including this essay's author, as said—upon whom the most severe supervision and control must be imposed.[18]

---

hideout with, among others, the MJH's incredibly rich library. (Another revelation of that time for me was that professional advisors—as a rule—do not read the same literature like me.).

[16] This concern was often propagated irrespective of the philosophical background of the commentators, cf., e.g., Stiglitz (1999) and Pejovich (2005).

[17] Mainly in this context, we discussed the parallel developments in East Germany (DDR). Unlike at the Holler's it was not always easy in Germany to drop my—then frequent—politically\incorrect comment that the Czechs were lucky not to have a rich brother in the West. (In this context it was noteworthy that our poorer brother Slovakia decided to turn back to our helping hand.) Much later I could extend upon this unification-disintegration debate during my advisory missions to North Korea (DPRK)—sponsored by the German government through the FDP's Friedrich Naumann Stiftung.

[18] And the same—often perverse—motivation can be ascribed to scholars who attack policies designed by their academic competitors and architect of the institutional change in the CEE could

3 Social Science for Practical Use And-Or its Own Sake

3.1 Theme No. 1: Convergence of an Auction

Having won the struggle over whether to privatize at all, what remained was the practical problem of how to do it in a country where the number of SOEs was in the order of thousands.[19]

Hence, what was needed was not just an "ordinary" but *mass* privatization. Our answer to this challenge was introduced under the label of a *voucher privatization.*

Any citizen over 18 years could register as a voucher-holder ("VH") and thus obtain 1000 investment points with the right to pay with them for shares of the companies into which the privatized SOEs were converted.

The exchange of shares for investment points was made in the form of a nationwide auction.[20] The idea about the auction's magnitude can be obtained from that in its Round 1.

On the supply side 300 million pieces of shares representing the book values of 1 491 privatized SOEs were simultaneously offered.

On the demand side8.54 million voucher-holders (some 70% of the adult population) represented the overall purchasing power of $8.54.10^9$ investment points (8.54 million *times* 1 000 points).

The role of an Auctioneer was assigned to a task force established by VK at his Federal Ministry of Finance. Given that the privatized shares were never traded, the Auctioneer could not have any notion about their market values. Hence, for Round 1 the reservation prices for all respective shares were set at the same level. The outcome of Round 1 thus had the form of mutually different excesses of demand or supply. This data for all 1 491 companies could be thus taken by the Auctioneer as the historically first relative valuation of the privatized SOEs.

The *over*-subscribed shares were all re-offered in Round 2—at prices increased in the proportion of the respective excess demands. By contrast, the *under*-subscribed shares were all sold to the respective VHs and the unsold shares were re-offered in Round 2 at prices proportionally decreased. In this *tatonnement* fashion, the auction rounds were to be repeated so that, ideally, all shares were sold, and all the investment points spent. Economically speaking, to arrive at this *zero outcome*, 1 491 equilibrium prices of the privatized shares had to be found out.

---

only speculate about the genuine motivation of the harsh attack upon the uneducated eastern natives by Stiglitz (2000). One—rather daring—answer could be that this brilliant academician turned Vice-president of the World Bank (IBRD) could not simply withstand the fact that some of us did not embrace his concepts of an information asymmetry and corporate governance as the most adequate means vis a vis the scope and scale of our real-world problems. (For a nice response to Stiglitz's proposals see Roháč 2013).

[19] In the 1980s, the British "case-by-case" privatization of Margaret Thatcher involved no more than tens of SOEs.

[20] For more details cf. Tříska (2009). A nice comparison between the Czech and Russian voucher auctions can be found in Boycko et al. (1995).

By nature, a privatization is always a political drama. In the case under study, the first dramatic ingredient was that the auction simply had to produce the *zero outcome*—anything else would equal to political disaster. Not only this, the outcome had to be achieved under a merciless time constraint brought up by the expected extinction of the seller—the Czech-Slovak Federal Republic. Based on the Auctioneer's intuitive guess of the date of the expected split, the Auctioneer came to the conclusion that there were no more than five auction rounds to fulfill this political mission.

And this is how the economics—practiced at the Liblice Conference—entered this otherwise fully political undertaking. Exactly there the participants were reminded that there exists a modern auction theory—apart from the Lange-Lerner proposals on how to do replace genuine markets.[21] Moreover, at Liblice, the Auctioneer was also enlightened by rather unorthodox applications of the game theory.[22] As a result, only a few months after Liblice the Auctioneer was able to interpret the privatization *tatonnement* as *a multi-player repeated game with a finite number of rounds* (five of them, in particular).

In fact, the privatization auction involved at least two kinds of games. The first was played among the VHs themselves as the outcome of every VH certain depended on the strategies of other VHs. The second game was between the Auctioneer (Player A) and the representative (aggregate) Voucher Holder (Player B).

Under the pressure to accelerate the auction's convergence, Player A included into the *rules of the game* the hypothesis of a *faithful voucher holder*. Player B thus learned that the Auctioneer will not defect from the strategy based on the assumption that VHs will be essentially faithful to their bids from the preceding round.[23]

However, despite all the scientific backing, no-one could guarantee the auction's timely convergence. What helped to get the process realized, was the mutual trust established at Liblice between the Auctioneer and the other two Conference participants, namely VK and MJH.

3.2 Theme No. 2: The True Nature of Central Planning

Let it be admitted that our strategy not to invite normative papers to the Liblice Conference was a mere extension of the policy we practiced in our Institute of Economics.[24]

---

[21] Cf. Larry (!) Samuelson (1990).

[22] Cf. Holler and Host (1990). It may be of value to note that Nobel Prizes are being awarded from time to time to both the auction and game theory.

[23] Data on the actual speed of the auction's convergence can be found in Tříska (2009). Of general interest could be that simultaneously with our time trial, the Czech prime minister VK held talks with his Slovak counterpart about the exact day "D" of the ultimate SLEXIT.

[24] Most CEE countries shared the model where economic institutes belonged to the respective Academy of Sciences and—as a rule—were relatively more liberal than, e.g., economics departments at universities.

To illustrate the Institute's non-normative production, we replaced the term *centrally-planned economy* with a value-free label *economy of a socialist type* (EST) and consecutively approached this EST as any other economy. By this we stressed that our subject-matter—EST - is different from "ordinary" market economies—by only a higher weight assigned to the command component of the coordination mechanism.[25]

In the same manner, our papers for Liblice argued that consumers and producers "under EST" can and must be analyzed as any other economic agents - that they also maximize their particular *preferences* subject to particular *constraints.*

Still more provocative could appear out claim that agents under EST must be classified as fully *rational*, however "peculiar" may appear their choices and behavior to outside observers. Our argument was that an agent under EST can be classified as *non-rational* only by scholars who will *insufficiently* specify the nature of preferences and constraints brought about by the particular institutional setting.[26]

Directly or indirectly, we thus strongly attacked the then popular proposals to conceive of EST as a "different animal" and analyze it by such highly normative methodological extravaganzas as "anti-equilibrium", "shortage", and "soft budget constraint".[27] To illustrate, we simply replaced the extravagant term "shortage" with a concept as ordinary as are the "search costs".[28]

In the light of the political economy of Sect. 2, the agreement on EST helped to clarify similarities and differences among *the starting points* of transformation strategies across the CEE.

3.3 Theme No. 3: Analytical Blocks for Inter-disciplinary Bridges

In his recent monograph[29] the author of this essay proposed to accept the notions of *preferences* and *constraints* as the *universal analytical building blocks* by which

---

[25] This concept fully extends upon Samuelson and Nordhaus (1985) in that: "No modern human economy is of a pure form. Rather, societies are mixed economies, with elements of market, command, and tradition. There has never been a 100% pure form."

[26] This observation we mainly ow to Stigler and Becker (1977), Posner (1993), etc.

[27] Cf. Kornai (1972, 1980, 1986). To our refusal of his methodological proposals he - in Kornai (2000a, 2000b) - mercilessly responded as follows: "… there are the errors and misdeeds committed by governments, officials, politicians, employers and employees, and parties and organizations. Dishonesty, corruption, negligence and incompetence … I find it understandable that the citizens of post-socialist countries do not carefully analyze and distinguish the separate causes just mentioned, and simply feel angered or embittered by the problems. However, what is understandable in lay citizens becomes unacceptable in social scientific researchers, highly qualified intellectuals, opinion-makers, and above all, politicians. These distinct groups cannot be allowed to get away even with well-meaning superficiality and ignorance, let alone with intentionally confusing the various causes of the country's problems …".

[28] Cf. the already quoted Tříska (1989).

[29] Tříska (2017).

social choice (decision-making) can and should be described-explained.[30] In addition, further universal blocks were based upon the trivial fact that a choice must be also analyzed with respect to whether and how exactly it is being executed. For that matter, we constituted two kinds of agents—a decision-maker (a so-called Designer) and a Designee whose *task* is to execute the Designer's choice.[31] In other words, a strict borderline was drawn between the "mental" category of a *choice* (decision making) and the "physical" category of a *behavior*, or the execution of the choice's outcome.

By calling the choice's outcome a *task* (*obligation*) it was stressed that what must be further analyzed is a hierarchy between a Designer and the subordinated Designee.

Whereas in legal scholarship the hierarchy is mostly established between different persons, in textbook economics a Designer is by and large assumed to design his-her own tasks—then rather called plans (goals, objectives, etc.).

Leaving aside terminological struggles, the proposal has been to accept the notion of a *task* (obligation, plan, goal, or objective) as another universal building block of a social scholarship. Finally, the notion of a task's *conditions* was accepted to complete the set of analytical blocks. Hence the notion of a behavior was to be described-explained in the format of conditions (states of the world) under which only the respective task can be prescribed for fulfillment.

To summarize, the proposal can be seen as a "yet another attempt" to construe an inter-disciplinary bridge: Whereas *preferences* and *constraints* are adopted from economics, a *task* and *conditions* find their roots in legal scholarship.[32] For want of a better term the bridge was termed General Theory of Choice and Behavior (GTCB).

To conclude a warning may be in order. Attempts such as the author's GTCB in fact contradict the generally shared concepts of inter- or multi-disciplinarity. Economics and legal scholarship are in fact taken as nothing more or less as two instances of the same GTCB.[33]

---

[30] The warning here could be that our approach to preferences and constraints may appear to be inconsistent with that in Buchanan (1966a).

[31] The fundamental analysis of the two roles is at length corroborated in Tříska (2017).

[32] In terms of the imperialistic ambitions of economics, we should speak about a bridge between economics and its scientific neighbors - cf. Buchanan (1966b).

[33] The relevance of GTCB's emergence to this Festschrift rests in that the actual work on Tříska (2017) effectively started in 2004 during the author's one-month self-imposed sabbatical at MJH's Hamburg University. Regardless how much or little was MJH was impressed by this academic project, the hard data demonstrate that he provided the author with a chair, table, and computer at Von-Melle-Park 5 (Hamburg University) and – as a bonus – an accommodation at his apartment at Kleiner Kielort, a walking distance from the office. On the top of all this, some 13 years later—shortly after the Opus Magnum Tříska (2017) was released, MJH and his Barbara came to Prague—by bicycles all the way from Munich—to make it possible for me to endow them with the book in person. The event is worth noting also because the book is printed in a hardcover, has almost 400 pages, and the two German heroes, let us recall, were to transport it to Munich by bicycles.

3.4 Theme No. 4: Rationality

By the term *agent* we will understand an object of a *social* origin—be it a business company, your neighbor John, a member state of the UN, the UN itself, etc. As already noted, we often regard an agent as *non-rational*, only because we wrongly identify his-her *preferences* and-or *constraints* or foolishly compare them with what we see as "objectively economic".[34]

By contrast, we propose a fully *subjectivist* concept according to which an agent can be classified as rational if he-she:

Axiom 1: believes to *understands* the nature and origins of his-her "problem",

Axiom 2: *can establish* a set of strategies about which he-she believes to be both relevant and *feasible*. The *infeasibility* of other relevant strategies thus constitute *constraints* to his-her particular purpose,

Axiom 3: *can compare* any pair of the strategies with respect to his-her *preferences*,

Axiom 4: *is able to select* one and only one strategy as his best/optimal strategy how to resolve his-her problem.

It is important to emphasize that this axiomatic system is based upon the agent's stubborn subjectivity. For a dramatic effect, let Axiom 1 be satisfied by John's clearly expressed fear of an invasion of Martians. Similarly, also "unrealistic" may appear John's belief (Axiom 2) that his problem may be relieved by leaving the Earth in the Air Force One. And so on with Axioms 3 and 4.

By contrast, we will classify John as non-rational, only if he cannot meet the above requirements, namely that of Axiom 4. Hence, he will be non-rational if he came to a conclusion that the planets Pluto and Jupiter are equally optimal (best) destinations for his Air Force One.[35]

For the sake of completeness, this subjectivism has been elsewhere applied also upon such concepts as efficiency, optimality, justice, etc.[36] And the same approach should be taken towards the very concept of the truth.[37]

---

[34] Cf. Knight (1982): "… the term economic has come to be used in a sense which is practically synonymous with intelligent or rational.".

[35] This particular case of a non-rationality is often demonstrated as the paradox of Buridan's ass who inevitably dies from hunger between two identical piles of hay. Another instance of a non-rationality would arise if the agent finds a still better solution to every solution. In medicine, breaches of our Axiom 4 are called *aboulomania*, a diagnosis representing a mental disorder in which the patient displays pathological indecisiveness.

[36] MJH—himself an expert on the subject—cf., e.g., Holler (1983)—never entered into a deeper discussion on rationality with me and would rather send me Chakravarty (2020) where the author comments Holler (2020).

[37] In this sense, Buchanan (1982) quotes Frank Knight as follows: "… truth is whatever emerged from the free discussion of reasonable men who approach the dialogue without prejudice and as good sports".

4 The Prisoner's Dilemma Revisited

4.1 A Social Economic Organization ("SEO")[38]

4.1.1 Reallocation of Valued Assets

It seems to be generally accepted that one of the major functions of an SEO is to reallocate valued assets among the SEO's members.[39]

Given that the interests of the members are—in principle—mutually different or even conflicting, the reallocation must be made through specific institutions, or social rules.[40] We will say that the actual design of an SEO's institutional setting is an outcome of a choice of an agent introduced above under the label of a Designer.[41]

4.1.2 A Voting System

For the sake of this analysis, an institutional setting of an SEO will be represented by a *voting system* through which its members may affect—by their votes—the resultant reallocation of assets.

Hence, an SEO's member will be taken—in addition to the Designer—as a second kind of a decision-maker.

4.1.3 A Judge

To simplify further, the Designer's choice of a *voting system* will be represented by specific *tasks* and *conditions* of an agent introduced above under the label of a Designee, namely his-her *tasks* to "process" the members' votes. Put in more detail, by design, a Designee will be obliged to

(a) collect and validate the member's votes,
(b) insert the votes into a *formula* designed for the particular purpose by the Designer,[42]

---

[38] The term represented by the acronym SEO is due to Knight (1986):

[39] Here we will use the terminology adopted from the economics of the so-called Edgeworth box by which is represented how an initial endowment of wealth is re-allocated between respective agents. In fact, a substantive part of Tříska (2017) is based upon this tool-king.

[40] According to Samuelson (1985): SEO's functions are defined by such questions as: what to produce, how to produce it and for whom? The social rules are then classified as traditions, commands, and market (free exchanges).

[41] We certainly use the term in the sense of the so-called "mechanism design models", see, e.g., Hurwitz (2007).

[42] Later, a so-called pay-off matrix will represent a formula into which votes are being inserted by the Judge.

(c) calculate out from the formula a *verdict* about the asset's reallocation.[43]

A Designee who—by design—has to fulfill the tasks will be called—for want of a better term—a Judge (a referee, a fact finder, etc.).

4.1.4 Who is Versus is not a Decision Maker?

It is tempting to conceive of the Judge's *verdict* as an outcome of his-her choice, or even a choice of an SEO itself.

Hence a warning is in order that there remains nothing to decide, once the *formula* is fixed by the Designer and the votes submitted.

Summarizing, then, once the two decision-makers make their choices, the remaining operations are nothing more or less than "mechanical calculations".[44]

4.2 The Case of a Plaintiff and a Defendant

4.2.1 A Collective Defendant

Invoking that our subject concerns Prisoners, it should be of no surprise that the *focal* (*blue-colored*) level of our further analysis will be constituted by an *SEO* consisting of a *Plaintiff* and a *Defendant*. (For technical reasons, *italic* underlined text represents "blue", while bold underlined text is used for "red").

Similarly expected should then be that an adjacent (red-colored) **SEO** will be introduced, namely by the assumption that the *Defendant* is a two-member ***gang*** consisting of a **Prisoner A** and a **Prisoner B**.[45]

4.2.2 Valued Assets to Be Reallocated

Let us summarize as follows:

(a) On the focal level of an *SEO*: whereas a *Plaintiff* seeks to maximize social justice, a *Defendant* wants to minimize punishment. To resolve this conflict of interests, the two members will use the *voting system* designed by a *Designer*.
(b) On the (red-colored) level of an **SEO**: the two **Prisoners** will evaluate in a conflicting mode variant reallocations of the punishment allocated to the

[43] The task of a Judge can be well illustrated by a parliamentary canvasser (scrutineer) who is to calculate out a parliament's choices by inserting into the respective formula data extracted from the MP's votes.

[44] The term "mechanical calculator" is frequently used by both Knight and Buchanan.

[45] In Holler (2018), a Judge is considered in the form of a 3-member jury. One should then ask who will be the Judge fulfilling the task to process the jurors' votes. This is how the mystery of an infinite recursion may complicate our debate. (Cf., e.g., the legendary question "*who will guard the guardians?*" in Hurwitz 2007).

Defendant on the focal level a). For that matter, **Prisoner A** and a **Prisoner B** will use the **voting system** designed by a **Designer**.

4.3 Comments

On the proviso that the analysis of preceding paragraphs is acceptable, we will dare briefly sketch a few directions how the Prisoner's Dilemma model could be revisited:

(1) A counterparty of the Prisoners—a Plaintiff, in our case—seems to have been left aside by the model—as if he-she should be indifferent to whatever may come out from the pay-off matrix. As a result, e.g., no inefficiency can be found in a reallocation where a decrease in punishment would not harm the *Plaintiff*'s concern for social justice. Moreover, an opportunity is thus missed to—by the intervention of a *Plaintiff*'s *vote*—prevent the unfortunate cases where the pay-off matrix provides an ambivalent solution or no solution at all.[46]

(2) The textbook model seems to suggest that the *verdict* made on the focal level of the *SEO* includes also the "lower-level" **verdict** on the reallocation of the overall punishment between the two Prisoners, i.e., a **verdict** belonging to the adjacent level of the **SEO**. Firstly, we should ask whether this mixture of the two analytical levels is correct. Secondly, in thus "aggregated" setting the two Prisoners will want to—by their "lower-level" **votes**—design the *Defendant*'s "focal-level" *vote*. To carefully interpret a setting where agents vote how somebody else should vote will certainly deserve a separate debate.

(3) An institutional setting that does not yield a unique outcome of a choice is evidently, if not "objectively" wrong in the sense that it contradicts the trivial fact that agents always somehow decide.[47] It is then only natural to expect that it will be a Designer's *task* to redesign any wrongful institutional setting. If so, the Designer of the Designer's *task* can personify the answer to the question in whose interest it may ever be to issue value judgments over variant reallocations of *somebody else's* valued assets.[48]

4) It is noteworthy that a Designer makes his-her choice under full uncertainty about how the SEO's members will vote. Our point here is that it is exactly this uncertainty that "upgrades" a Designer into the role of a genuine *problem-solver*. By contrast, a Judge, who is—by assumption—fully informed about the votes, will be nothing more than a "mechanical calculator" of a verdict.[49]

---

[46] Typically for MJH, he informed me about Binmore (2021) who seeks to show "… how conventions operate to select a Nash equilibrium from a game with many Nash equilibria …".

[47] To avoid this contradiction, e.g., non-convex preferences are refused to represent a real-world consumer.

[48] MJH invited me in May 2016 to his Adam Smith Seminar to deliver a talk on "pseudo-normativity of economics" where one of my ambitions was to refuse Pareto-efficiency as an *objective* criterion of choice.

[49] If we quote, again, from the MJH's gift to the author of the essay, in Buchanan (1966a) argues as follows "*If I know what I want, a computer can make all of my choices for me.*"

To conclude, MJH will be reminded of the author's frequently repeated question: "*Why "on earth" the Prisoner's choices (to confess or deny) should deserve any special treatment? In what way they should be substantially different from other choices under uncertainty?*". (The author's arguments have always been that once a choice of somebody else is admitted to the model as its exogeneous parameter, it simply must be of the same nature as, e.g., a change of weather or an outbreak of war.)

5 Summary and Conclusions

The events, topics, and theses selected for this essay should primarily illustrate the uniqueness of the author's relationship with Prof. Manfred J. Holler—ever since it emerged from the (legendary) Liblice Castle Conference (March 1989, then Czechoslovakia).

Within the *normative* part of the essay, the author's approach to the relationship between *power* and *responsibility* is by and large based on his political assignments from the 1990s, i.e., the time when his country undertook its transition from the *real* socialism to a would-be *democratic* capitalism. His notes and comments deal namely with:

- the sensitivity of trade-offs between conflicting political objectives, namely, those of an economic and a non-economic nature,
- often dubious motivations of scholars turned politicians.

In the realm of a *positive*, value-free science, the author:

- attempts to propagate his mono-disciplinary proposal to take all the societal disciplines as *mere* instances of one single General Theory of Choice and Behavior,
- introduces his extensively subjectivist concept of rationality where non-rationality is in fact *only* a diagnosis of a specific mental disorder called *pathological indecisiveness*,
- uses the Prisoner's dilemma model to demonstrate the variety of levels on which a societal problem can and should be analyzed and that the borders between the levels are not always strictly respected.

References

Binmore, K. (2021). Thinking about social norms. Paper prepared for Adam Smith Seminar 2021

Boycko, M., Shleifer, A., Vishny, R. (1995). Privatizing Russia. Massachusetts Institute of Technology. ISBN 0–262–52228–4

Buchanan, J. M. (1964). What should economists do?. *Southern Economic Journal*. *30*(3), 213–222

Buchanan, J. M. (1966a). *What Should Economists Do?*. Indianapolis: Liberty Fund. ISBN 978–0–913966–65–5
Buchanan, J. M. (1966b). *Is Economics the Science of Choice?* (pp. 39−54). In BUCHANAN (1966b)
Buchanan, J. M. (1966c). *Economics and Its Scientific Neighbors* (pp. 115–142). In: BUCHANAN (1966c)
Buchanan, J. M. (1982). Preface to KNIGHT (1982).
Buchanan, 1993Buchanan, J. M. (1993). *Property as a Guarantor of Liberty*. The Locke Institute, Edward Elgar. ISBN 1852787333
Burgin, A. (2009). The radical conservatism of frank H. Knight. *Modern Intellectual History*, *6*(3), 513–538
Chakravarty, S. P. (2020). Ambiguity of Rationality. Comments to HOLLER, J. M. (2020), prepared for Adam Smith Seminar 2021
Coase, R. H. (1999). Economists and Public Policy, In KLEIN (1999).
Friedman, M. (1953).*The Methodology of Positive Economics*, *Essays in Positive Economics* (pp. 3–43). University of Chicago Press
Gravelle, H. & Rees, R. (2004). *Microeconomics* (3rd ed). Pearson Education Canada. ISBN 10 0–582–40 187–8
Hayek F. A. (1980). *Counter-Revolution of Science*. Liberty Fund. ISBN 978–0913966679
Hayek F. A. (1991). The Fatal Conceit. In W. W. Bartley (Ed.), *The Errors of Socialism* (The Collected Works of F. A. Hayek). University of Chicago Press. ISBN-10: 0226320669
Holcombe, G. R. (2015). Political capitalism. *Cato Journal*, *35*(1), 41–66
Holler, J. M. (1983). Collective action, rational man and economic reasoning. *Quality and Quantity*, *17*, 163–177. Elsevier Science Publishers B.V., Amsterdam -- Printed in The Netherlands
Holler, J. M. & Host, V. (1990). Maximin vs Nash Equilibrium: Theoretical Results and Empirical Evidence. In R. Quandt, & D. Triska
Holler, J. M. (2018). Marquis de Condorcet and the Two-dimensional Jury Model. In A. Marciano, & G. Ramello, (Eds.), *Law and Economics in Europe and the U.S*. The Legacy of Jürgen Backhaus
Holler, J. M. (2020). *The Economics of the Good, the Bad and the Ugly. Secrets, Desires, and Second-Mover Advantages*. Routledge. ISBN 13: 978–0367587772
Hurwicz, L. (2007). But who will guard the guardians?. *Nobel Memorial Lecture*. available at w
Kaplow, L., & Shavell, S. (2009). *Fairness versus Welfare*. Harvard University Press. ISBN 0–67403–931–9
Klaus, V., & Tříska, D. (1995). Recenzní zamyšlení nad knihou Jánose Kornaie "Socialistický systém: Politická ekonomie komunismu". (*On János Kornai's "Socialist System: The Political Economy of Communism*), Finance a úvěr, č. 3/1995, str. 137 – 142.
Klaus, V., & Tříska, D. (2006). Polemika se Svetozarem Pejovichem o transformační, tedy neklasické privatizaci. (*A Polemic with Svetozar Pejovich on the Transformational, i.e. Non-classical Privatization*"). Politická ekonomie, vol. LIV, No. 3/2006, pp. 291–306. English abstract available at http://www.vse.cz/polek/abstrakt.php3?IDcl=559
Knight, F. H. (1986). Social Economic Organization. In W. Breit, H. M. Hochman, & E. Saueracker, (Eds.), *Readings in Microeconomics* (pp. 1–13). Times Mirror/Mosby College Publishing. ISBN 0–8016–0795–7
Knight, F. H. (1982a). *Freedom and reform. Essays in Economics and Social Philosophy* (1st ed.). Liberty fund. ISBN 0–86597–004–1
Knight, F. H. (1982b). Freedom as Fact and Criterion, Chapter I. In F. H. Knight. (pp 3 – 23).
Kornai, J. (1972). *Anti-Equilibrium Amsterdam*. New York: American Elsevier Publishing Company, Oxford, North-Holland, Inc. 0-720430-550
Kornai, J. (1980). *Economics of Shortage*. Elsevier North-Holland. ISBN 0–44486–059–2
Kornai, J. (1986). The soft budget constraint. *Kyklos, 39*(1986), 3–30.
Kornai, J. (2000a). What the change of system from socialism to capitalism does and does not mean. *Journal of Economic Perspectives*, *14*(1), 27–42

Kornai, J. (2000b). *The Socialist System. The Political Economy of Communism*. Princeton, NJ: Princeton University Press, and Oxford: Oxford University Press. ISBN-0–691–00–393–9
Klein, 1999Klein, D.B. (Eds). (1999). *What Do Economists Contribute?*. New York: University Press. 0-8147-4723-X
Okun, A. M. (1975). *Equality and efficiency; The big tradeoff*. The Brooking Institution. ISBN 0 8157 6475 8
Ostrom, E. (2006). *Understanding institutional diversity*. Princeton University Press. ISBN 978–0691122380
Ostrom, E. (2009). *Beyond markets and states: Polycentric governance of complex economic systems*. Prize Lecture
Pejovich, S. (2005). On the privatization of "stolen goods" in central and Eastern Europe. *The Independent Review*, *10*(2). ISSN 1086–1653 209 229
Phelps, E. (2015). What is wrong with the west's economies. *New York Review of Books*. *62*(13)
Posner, R. (1993). What do judges and justices maximize? (The Same Thing Everybody Else Does). *Supreme Court Economic Review*, *3*, 1–41
Quandt and Tříska, 1990Quandt, R., & Tříska, D. (Eds). (1990) *Optimal decisions in markets and planned economies*. London: Westview Press. ISBN 0-8133-0994-8
Rudiger, F., & Burghart, S. (Eds.). (2009): *Driving forces of socialist transformation; North Korea and the experience of Europe and East Asia*. Wien: Praesens Verlag. ISBN 978–3–7069–0569–5
Rohac, D. (2013). What are the lessons from post-communist transitions? *Economic Affairs, Wiley Blackwell, 33*(1), 65–77.
Samuelson, L. (1990). Optimal Auctions and Market Structure. In R. Quandt, & D. Triska
Samuelson, P. A., & Nordhaus, W.D. (1985). *Economics* (12th ed.). McGraw-Hill. ISBN-0–07-Y66533–8
Shelling, T. C. (1999). What Do Economists Know?. In Klein
Stigler, G. J., & Becker, G. S. (1977). De Gustibus Non Est Disputandum. *The American Economic Review*, *67*(2), 76–90
Stiglitz, J.E. (1999). *Whither reform? ten years of the transition. keynote address to the annual world bank conference on development Economics*. Washington: D.C. April 28–30
Tříska, D. (1989). Consumer under supply constraint. *European Journal of Political Economy*, *5*, Elsevier Science Publishers B.V
Tříska, D. (2005). K některým možnostem optimalizace smlouvy a závazkových vztahů. (Towards Optimization of Contracts and Obligation Relationships), Oeconomica, Praha 2005. ISBN 80–245–0916–4
Tříska, D. (2009). Methods of privatization in western, central and Eastern Europe. In F. Rudiger, & S. Burghart (pp. 159–187)
Tříska, D. (2016). A comment on PHELPS (2015). *Homo Oeconomicus Journal of Behavioral and Institutional Economics*. *33*, 1–2. ISSN 0943–0180
Tříska, D. (2017). *Social choice and behavior; the nature of their design and management*. Praha: Walters Kluwer. ISBN 978–8–07552921–3. http://www.fsp.cz/social-choice-and-behavior/

Lexicographic Ranking Based on Minimal Winning Coalitions

M. Aleandri, V. Fragnelli, and S. Moretti

1 Introduction

The *desirability relation* for players in simple games Carreras and Freixas (1995); Freixas et al. (2012); Isbell (1958); Taylor and Zwicker (1999) has been widely studied also in connection with the property-driven analysis of power indices Diffo Lambo and Moulen (2002); Freixas and Gambarelli (1997); Holler and Nurmi (2013). A player i is in desirability relation with a player j (meaning that i is at least as desirable as j) if we can replace player j with player i in any winning coalition without changing the outcome, i.e. for any winning coalition S with $j \in S$ and $i \notin S$, we have that $S \setminus \{j\} \cup \{i\}$ is still a winning coalition. So, the desirability relation between i and j suggests that player i is at least as influential as player j, for it is never harmful for coalitions to replace j by i. Clearly, the desirability relation is not necessarily a total relation on the set of players, as two players may not be in any desirability relation (see, for example, Freixas and Pons (2005, 2008); Holler and Nurmi (2013) for an analysis of properties of the desirability relation on simple games). When the desirability relation is a total preorder the simple game is called complete and, for example, weighted games are complete, see Alonso-Meijide and Freixas (2010).

A classical property for power indices based on the desirability relation, is the monotonicity property: a power index is monotone if, whenever a player i is at least as desirable as player j, then the power of i is at least as much as the power of player j (see Remark 1). For instance, the Shapley-Shubik index Shapley and Shubik (1954), the Banzhaf index Banzhaf (1965), the Johnston index Johnston (1978), the

M. Aleandri
Luiss University, Viale Romania, 32, 00197 Rome, Italy
e-mail: maleandri@luiss.it

V. Fragnelli (✉)
Department of Sciences and Innovative Technologies (DISIT), University of Eastern Piedmont, Viale T. Michel 11, 15121 Alessandria, Italy
e-mail: franco.fragnelli@gmail.com

S. Moretti
LAMSADE, CNRS, Université Paris-Dauphine, Université PSL, 75016 Paris, France
e-mail: stefano.moretti@dauphine.fr

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_13

nucleolus Schmeidler (1969) and many other power indices satisfy the monotonicity property (see, for instance, Freixas and Gambarelli (1997)). Instead, it is easy to provide examples of simple games showing this property is not satisfied by other famous indices that take into account exclusively minimal winning coalitions, like the Deegan-Packel index (DPI) Deegan and Packel (1980) or the Public Good Index (PGI) Holler (1982); Holler and Packel (1983) (see Sect. 3 for some examples from the literature Deegan and Packel (1980); Freixas and Gambarelli (1997)).

The main objective of this paper is to show that it is possible to rank players consistently with the desirability relation and using exclusively minimal winning coalitions in an ordinal way. In fact, while classical power indices convert the information about coalitions into a numerical personal score representing players' relevance in a simple game, in many practical situations, having a reliable ranking to select the top players is enough and the information provided by players' score is only marginal. For instance, in the application of power indices to computational biology, the goal is short-listing the most relevant genes on complex networks with a huge number of nodes Moretti et al. (2007). In a similar way, ranking players are essential for the analysis of centrality of network elements with the goal to select the most critical or sensible parts of a system Lindelauf et al. (2013), or in studies aimed at establishing which agents are the strongest or the weakest in a voting system Fertö et al. (2020).

To that purpose, we introduce a *ranking solution* (formally, a map that associates to any simple game with player-set N, a total preorder on N) aimed at ranking players in a simple game according to their influence and in a way that is compatible with the desirability relation. Our ranking solution contains elements of both the DPI and the PGI, taking into account the minimal winning coalitions an individual belongs to. More exactly, given $n = |N|$ players, we first compute for each player a vector of n real numbers, where the k-th component of each vector is the number of minimal winning coalitions including the player of size k, with $k = 1, \ldots, n$; second, our ranking solution lexicographically compares those real-valued vectors. Due to the similarity with the PGI to compute vectors components, we called such a ranking solution the *Lexicographic Ranking based on Minimal winning coalitions* (shortly, the *LRM*).

We show that the *LRM* is monotonic, and we prove that it is the unique solution satisfying (strong) monotonicity with respect to the desirability relation together with two other axioms: (1) the *coalitional anonymity* property, saying that the relative ranking between two players i and j in two different simple games should be independent of the identity of other players in minimal winning coalitions, provided that the number of minimal winning coalitions to which they belong in the two games is the same; (2) the property of *independence of larger minimal winning coalitions*, saying that once a player i is considered more influential than a player j in a simple game, player i will continue to be considered more influential than j in any simple game obtained by adding new "larger" minimal winning coalitions to the original game.

As a side-product of our analysis, we also point out some connections between our ranking solution and the criticality-based ranking provided in Aleandri et al. (2021) to compare, in an ordinal way, the blocking power of players and we explore some

similarities with the axioms used to characterize the dual version of the criticality-based ranking.

The paper is structured as follows. In Sect. 2 we provide some basic notions and notation. The definition of the *LRM* is then introduced in Sect. 3 together with some examples comparing it with the ranking defined by other solutions from the literature. An axiomatic characterization of the *LRM* is then presented and discussed in Sect. 4. A connection between the criticality-based ranking and the *LRM* is investigated in Sect. 5 using the desirability relation on dual games. Section 6 concludes.

2 Preliminaries and Notation

Given a finite set N, we denote by $|N|$ its *cardinality* and by $2^N = \{S \subseteq N\}$ its *power set*. A *simple game* is a pair (N, v), where $N = \{1, 2, \ldots, n\}$ denotes a finite set of players and $v : 2^N \to \{0, 1\}$ is a *characteristic function*, with $v(\varnothing) = 0$, $v(N) = 1$ and $v(S) \leq v(T)$ for all sets S, T such that $S \subseteq T \subseteq N$. A coalition $S \subseteq N$ such that $v(S) = 0$ is said a *losing* coalition, whereas a coalition $S \subseteq N$ such that $v(S) = 1$ is said a *winning* coalition. The class of simple games with N as the set of players is denoted by $\mathcal{SG}^N$.

Let $\mathcal{W}^v$ be the set of winning coalitions in (N, v)

$$\mathcal{W}^v = \{S \subseteq N : v(S) = 1\}$$

and let $\mathcal{W}^v_{\min}$ be the set of minimal winning coalitions in (N, v)

$$\mathcal{W}^v_{\min} = \text{Min } \mathcal{W}^v$$

where, for any family of sets $\mathcal{F}$, the Min operator on $\mathcal{F}$ removes all non-inclusion-minimal sets of $\mathcal{F}$:

$$\text{Min } \mathcal{F} = \{F \in \mathcal{F} | \nexists G \in \mathcal{F} : G \subset F\}.$$

A simple game (N, v) is a *weighted majority game* if there exists a vector of non-negative real numbers $w \in \mathbb{R}^N_{\geq 0}$ and a quota $q \in \mathbb{R}_{\geq 0}$ such that a coalition $S \subseteq N$ is winning if and only if $\sum_{i \in S} w_i \geq q$.

In Holler (1982) the author introduced the *Public Good index* (*PGI*) of a player in a simple game, as the quotient between the number of minimal winning coalitions containing that player and the sum of cardinalities of all the minimal winning coalitions. Let (N, v) be a simple game, the PGI of player $i \in N$:

$$h_v(i) = \frac{|\mathcal{W}^v_{\min}(i)|}{\sum_{j \in N} |\mathcal{W}^v_{\min}(j)|}$$

where $\mathcal{W}^v_{\min}(i) = \{W \in \mathcal{W}^v_{\min} : i \in W\}$.

Example 1 Let $N = \{1, 2, 3, 4, 5\}$, and let

$$\mathcal{W}^v_{\min} = \{\{1, 2\}, \{1, 3\}, \{3, 4\}, \{2, 4, 5\}, \{1, 4, 5\}\}.$$

We have

$$h_v(1) = \frac{3}{12},\ h_v(2) = \frac{2}{12},\ h_v(3) = \frac{2}{12},\ h_v(4) = \frac{3}{12},\ h_v(5) = \frac{2}{12}.$$

In Example 1 we can observe that player 3 and 5 have the same PGI, but they belong to minimal winning coalitions of different cardinality and whenever player 5 is winning together with some coalition $S \subseteq N \setminus \{3, 5\}$ then player 3 is winning together with the same coalition.

In Deegan and Packel (1980) the authors measure the power of a player according to the size of the minimal winning coalitions she belongs to. So, the *Deegan-Packel index* (*DPI*) for player i is defined as:

$$\delta_v(i) = \sum_{W \in \mathcal{W}^v_{\min}(i)} \frac{1}{|\mathcal{W}^v_{\min}|} \frac{1}{|W|}.$$

Example 2 Taking the same simple game (N, v) of Example 1 we have:

$$\delta_v(1) = \frac{8}{30},\ \delta_v(2) = \frac{5}{30},\ \delta_v(3) = \frac{6}{30},\ \delta_v(4) = \frac{7}{30},\ \delta_v(5) = \frac{4}{30}.$$

According to the DPI, player 3, for instance, has more power than player 5 because it belongs to two minimal winning coalitions of size smaller than the two minimal winning coalitions containing player 5.

3 A Ranking Solution and the Desirability Relation

Let us start recalling that a binary relation on N is a subset of $N \times N$. A reflexive, transitive and total binary relation on N is a *total preorder* (also called, a *ranking*) on N. We denote by $\mathcal{T}^N$ the set of all total preorders on N. For instance, consider the lexicographic total preorder among vectors of real numbers:

$$\mathbf{x} \geq_L \mathbf{y} \quad \text{if either } \mathbf{x} = \mathbf{y} \quad \text{or} \quad \exists k : x_t = y_t, t = 1, \ldots, k-1 \ \text{ and } x_k > y_k.$$

We define a *ranking solution* or, simply, a *solution*, as a map $R : \mathcal{SG}^N \to \mathcal{T}^N$ that associates to each simple game $v \in \mathcal{SG}^N$ a total preorder on N. The value assumed by a map R on a simple game v is the ranking on N denoted by R^v. We use the

notation $i\,R^v\,j$ to say that $(i,\,j) \in R^v$, and it means that i is at least as important as j according to ranking R^v, for all $i,\,j \in N$. We denote by I^v the symmetric part of R^v, i.e. $i\,I^v\,j$ means that $(i,\,j) \in R^v$ and $(j,\,i) \in R^v$ (i and j are equivalent), and by P^v its asymmetric part, i.e. $i\,P^v\,j$ means that $(i,\,j) \in R^v$ and $(j,\,i) \notin R^v$ (i is strictly more important than j).

Clearly, any real-valued N-vector numerically represents a total preorder over the player set N. Consequently, any power index $\phi : \mathcal{SG}^N \to \mathbb{R}^N$ underpins a ranking solution denoted by R_ϕ and such that $i\,R^v_\phi\,j \Leftrightarrow \phi_i(v) \geq \phi_j(v)$.

In this section, we introduce a new ranking solution for simple games based on minimal winning coalitions. The main idea of the new solution is that the smaller is the size of a minimal winning coalition, the larger is the power of its members. Therefore, the ranking of a player is positively correlated first to the size of minimal winning coalitions the player belongs to and, second, to their number.

To define the ranking solution, we need to introduce the notation i_k representing the number of minimal winning coalitions of size k containing i in a simple game (N, v): $i_k = |\{S \in \mathcal{W}^v_{\min} : i \in S, |S| = k\}|$ for all $k \in \{1, \ldots, n\}$. For each $i \in N$, let $\theta_v(i)$ be the n-dimensional vector $\theta_v(i) = (i_1, \ldots, i_n)$ associated to v.

Definition 1 [Lexicographic Ranking based on Minimal winning coalitions (LRM)] The *Lexicographic Ranking based on Minimal winning coalitions (LRM)* solution is the function $R_l : \mathcal{SG}^N \longrightarrow \mathcal{T}^N$ defined for any simple game $v \in \mathcal{SG}^N$ as

$$i\ R^v_l\ j \qquad \text{if} \qquad \theta_v(i)\ \ \geq_L\ \ \theta_v(j).$$

Let I^v_l and P^v_l be the symmetric part and the asymmetric part of R^v_l, respectively.

Example 3 Consider the simple game of Example 1. We have that

$$\begin{aligned}
&\theta_v(1) = (0, 2, 1, 0, 0),\ \ \theta_v(2) = (0, 1, 1, 0, 0),\\
&\theta_v(3) = (0, 2, 0, 0, 0),\ \ \theta_v(4) = (0, 1, 2, 0, 0),\\
&\theta_v(5) = (0, 0, 2, 0, 0).
\end{aligned}$$

So, the LRM solution R^v_l ranks the players as follows

$$1\ P^v_l\ 3\ P^v_l\ 4\ P^v_l\ 2\ P^v_l\ 5.$$

Notice that the ranking provided by the PGI and the DPI do not coincide with the ranking R^v_l on this example. In fact, for instance, $h_v(4) > h_v(3)$ and $\delta_v(4) > \delta_v(3)$, while $3\ P^v_l\ 4$.

The LRM solution always provides a total preorder over the player set N for any simple game (N, v). Instead, given a simple game (N, v), the desirability relation Isbell (1958) is a preorder over the elements of N and is defined as follows.

Definition 2 Let (N, v) be a simple game. For any pair of players $i, j \in N$, the *desirability relation* $\succeq^v \subseteq N \times N$ is defined as follows:

$$i \succeq^v j \Leftrightarrow [S \cup \{j\} \in \mathcal{W}^v \Rightarrow S \cup \{i\} \in \mathcal{W}^v \text{ for all } S \subseteq N \setminus \{i, j\}].$$

In the following, if the game v on which $\succeq^v$ is defined is clear from the context, we denote relation $\succeq^v$ simply by $\succeq$. For any $i, j \in N$, $i \succeq j$ is interpreted as player i is at least as desirable as player j (as a coalitional member); $i \succ j$ means that $i \succeq j$ and there exists a coalition $T \subseteq N \setminus \{i, j\}$ such that $T \cup \{i\} \in \mathcal{W}^v$ but $T \cup \{j\} \notin \mathcal{W}^v$, and it is interpreted as player i is (strictly) more desirable than player j; $i \sim j$ means that $i \succeq j$ and $j \succeq i$, i.e. it is true that $S \cup \{j\} \in \mathcal{W}^v \Leftrightarrow S \cup \{i\} \in \mathcal{W}^v$ for all $S \subseteq N \setminus \{i, j\}$, and it is interpreted as players i and j are equally desirable.

As discussed in Sect. 1, the desirability relation, when it holds, represents a criterion to select between two players the most influential one, that is the player winning the maximum number of times. So, it is interesting to require the following basic property for ranking solutions.

Property 1 *(Desirable Monotonicity (DM)) Let $i, j \in N$. For any $v \in \mathcal{SG}^N$, a solution R satisfies the desirable monotonicity property if*

$$i \sim^v j \Rightarrow i I^v j,$$

and

$$i \succ^v j \Rightarrow i P^v j.$$

A solution satisfying the desirable monotonicity should strictly obey to the desirability relation: if the desirability relation between two players is strict (i.e., $i \succ^v j$), then a ranking solution should put such players in a strict relation too (i.e., $i P^v j$); of course, if two players are equally desirable (i.e., $i \sim^v j$) then the ranking solution must define the same kind of relation (i.e., $i I^v j$). Notice that this kind of "strong" monotonicity relation is not satisfied by the ranking over players represented by the nucleolus Schmeidler (1969), as it is easy to find examples of simple games having players in the symmetric part of the desirable relation and such that the allocation provided by the nucleolus is different (see, for instance, Freixas and Gambarelli (1997) page 600).

The rankings over players represented by the DPI and the PGI do not satisfy the desirable monotonicity property, as shown by the following example.

Example 4 Consider a weighted majority game (N, v), $N = \{1, 2, 3, 4, 5\}$, with weight function $(4, 2, 1, 1, 1)$ and quota $q = 6$. So, the minimal winning coalitions are

$$\mathcal{W}^v_{\min} = \{\{1, 2\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 4, 5\}\}.$$

We have that $1 \succ^v 2 \succ^v 3 \sim^v 4 \sim^v 5$. However, according to the PGI $h_2(v) = \frac{1}{11} < \frac{2}{11} = h_3(v)$, while according to the DPI we have $\delta_2(v) = \frac{1}{8} < \frac{1}{6} = \delta_3(v)$. So, according to the rankings underpinned by both indices, player 3 is ranked strictly higher than player 2.

On the other hand,

$$\theta_v(1) = (0, 1, 3, 0, 0), \quad \theta_v(2) = (0, 1, 0, 0, 0),$$
$$\theta_v(3) = (0, 0, 2, 0, 0), \quad \theta_v(4) = (0, 0, 2, 0, 0),$$
$$\theta_v(5) = (0, 0, 2, 0, 0).$$

So, 1 P_l^v 2 P_l^v 3 I_l^v 4 I_l^v 5: R_l^v and $\succeq^v$ coincide.

In general, a total preorder provided by the LRM solution coincides with the desirability relation on any simple game where the desirability relation is total. This fact is an immediate consequence of the following proposition.

Proposition 1 *The LRM solution R_l fulfils the desirable monotonicity property.*

Proof Let (N, v) be a simple game. It is easy to verify that the condition

$$S \cup \{j\} \in \mathcal{W}^v \Leftrightarrow S \cup \{i\} \in \mathcal{W}^v \text{ for all } S \subseteq N \setminus \{i, j\} \tag{1}$$

is equivalent to the condition

$$S \cup \{j\} \in \mathcal{W}^v_{\min} \Leftrightarrow S \cup \{i\} \in \mathcal{W}^v_{\min} \text{ for all } S \subseteq N \setminus \{i, j\}. \tag{2}$$

We prove that $i \sim^v j \Rightarrow i\, I_l^v\, j$.

Since $i \sim^v j$, according to the equivalence between relations (1) and (2), we immediately have that $S \cup \{j\} \in \mathcal{W}^v_{\min} \Leftrightarrow S \cup \{i\} \in \mathcal{W}^v_{\min}$ for all $S \subseteq N \setminus \{i, j\}$. So, $\theta_v(i) = \theta_v(j)$, and therefore $i\, I_l^v\, j$.

Now, we prove that $i \succ^v j \Rightarrow i\, P_l^v\, j$.

Let be $i \succ^v j$ and define $\mathcal{T} = \{T \subseteq N \setminus \{i, j\} : T \cup \{i\} \in \mathcal{W}^v_{\min}, T \cup \{j\} \notin \mathcal{W}^v_{\min}\}$.

We first need to prove that $\mathcal{T} \neq \emptyset$. Since $i \succ^v j$, it must exist $T \subseteq N \setminus \{i, j\}$ such that $T \cup \{i\} \in \mathcal{W}^v$ and $T \cup \{j\} \notin \mathcal{W}^v$ and, by the equivalence between relation (1) and (2), it is not possible that $S \cup \{j\} \in \mathcal{W}^v_{\min} \Leftrightarrow S \cup \{i\} \in \mathcal{W}^v_{\min}$ for all $S \subseteq N \setminus \{i, j\}$. Moreover, again for $i \succ^v j$, it is not possible that there exists $S \subseteq N \setminus \{i, j\}$ such that $S \cup \{i\} \notin \mathcal{W}^v_{\min}$ and $S \cup \{j\} \in \mathcal{W}^v_{\min}$. So, it must exist $T \subseteq N \setminus \{i, j\}$ such that $T \cup \{i\} \in \mathcal{W}^v_{\min}$ and $T \cup \{j\} \notin \mathcal{W}^v_{\min}$.

Now, let $k = \min\{|T| : T \in \mathcal{T}\}$. If $k = 0$, we immediately have that $\{i\} \in \mathcal{W}^v_{\min}$ and $\{j\} \notin \mathcal{W}^v_{\min}$, so $i\, P_l^v\, j$.

Consider the case $k > 0$. By the minimality of k we have that $i_t = j_t$ for all $t = 0, \ldots, k-1$ and $i_k > j_k$ and so $i\, P_l^v\, j$. □

Remark 1 It is well known from the literature that the desirability relation on weighted majority games is a total preorder and that the following monotonicity condition w.r.t. weights holds for a weighted majority game (N, v) with weights $(w_1, \ldots, w_n)$:

$$w_i \geq w_j \Rightarrow i \succeq^v j,$$

for all $i, j \in N$ (see for instance Freixas and Gambarelli (1997)). As a direct consequence of Proposition 1 we have that also the LRM solution on weighted majority games is monotonic w.r.t. weights, that is $w_i \geq w_j \Rightarrow i R_l^v j$ for all $i, j \in N$.

4 An Axiomatic Characterization of the LRM Solution

Now, we introduce two new properties for ranking solutions that are inspired by similar properties introduced in Aleandri et al. (2021) on the sets of blocking coalitions.

The next property says that winning coalitions of the same size should have the same impact on the ranking, independently of their members.

Property 2 *(Anonymity of Minimal Winning Coalitions (AMWC))* Let $i, j \in N$, $v, v_\pi \in \mathcal{SG}^N$ and let π be a bijection on $2^{N\setminus\{i,j\}}$ with $|\pi(S)| = |S|$ and such that

$$S \cup \{i\} \in \mathcal{W}_{\min}^v \Leftrightarrow S \cup \{i\} \in \mathcal{W}_{\min}^{v_\pi}$$

and

$$S \cup \{j\} \in \mathcal{W}_{\min}^v \Leftrightarrow \pi(S) \cup \{j\} \in \mathcal{W}_{\min}^{v_\pi},$$

for all $S \in 2^{N\setminus\{i,j\}}$. A solution R satisfies the *anonymity of minimal winning coalitions* property if

$$i R^v j \Leftrightarrow i R^{v_\pi} j.$$

Example 5 Consider the weighted majority game (N, v) of Example 4 and the players 3 and 4 in the role of players i and j of the definition of Property 2. Define a bijection π on $2^{\{1,2,5\}}$ such that $\pi(\{1, 5\}) = \{2, 5\}$. So the simple game (N, v_π) is such that

$$\mathcal{W}_{\min}^{v_\pi} = \{\{1, 2\}, \{1, 3, 4\}, \{1, 3, 5\}, \{2, 4, 5\}\}.$$

Game v differs from v_π in terms of minimal winning coalitions just for coalition $\{1, 4, 5\}$ which is replaced in v_π by the minimal winning coalition $\{2, 4, 5\}$. Nevertheless, the number of minimal winning coalitions of each size containing player 4 in game in v_π is precisely as in game in v, so her capacity to form minimal winning coalitions should not be affected (assuming that the other players are equally inclined to form minimal winning coalitions with 4). So, the property of Anonymity of Minimal Winning Coalitions says that the relative ranking between 3 and 4 in v should be the same as in v_π.

Property 2 reflects a broadly adopted principle, satisfied by classical power indices like the Shapley-Shubik index Shapley and Shubik (1954), the Banzhaf index Banzhaf (1965) and all semivalues Dubey et al. (1981), saying that coalitions of the same size are equally likely. So, it seems compelling to assume that the relative

position of two players is not affected by permutations preserving the size of minimal winning coalitions containing them, as it is required by Property 2.

Another property we consider in our analysis is the one of independence of larger minimal winning coalitions, saying that, once a solution exists, in which a player i is ranked strictly better than a player j, adding "larger" minimal winning coalitions should not affect the relative ranking between i and j.

Property 3 *(Independence of Larger Minimal Winning Coalitions (ILMWC))* Let $i, j \in N$. For any $v \in \mathcal{SG}^N$, let $h = \max\{|S| : S \in \mathcal{W}^v_{\min}$ and $S \cap \{i, j\} \neq \varnothing\}$ be the highest cardinality of coalitions in the set $\mathcal{W}^v_{\min}$ containing either i or j. Let $\mathcal{S}_h$ be a collection of (minimal) winning coalitions with cardinality strictly larger than h, i.e., $\mathcal{S}_h = \{S_1, \ldots, S_r\}$ such that $S_k \subseteq N$, $|S_k| > h$ for $k = 1, \ldots, r$ and there is no $Q \in \mathcal{W}^v_{\min} \cup \mathcal{S}_h$ with $Q \subset S_k$, for all $k \in \{1, \ldots, r\}$. A solution R satisfies the *independence of larger minimal winning coalitions* property if

$$i P^v j \Rightarrow i P^{v'} j,$$

where v' is a simple game such that the set of minimal winning coalitions is obtained as $\mathcal{W}^{v'}_{\min} = \mathcal{W}^v_{\min} \cup \mathcal{S}_h$.

Example 6 Consider again the weighted majority game (N, v) of Example 4 and the player 1 and 2 in the role of players i and j of the definition of Property 3. Let $\mathcal{S}_h = \{\{2, 3, 4, 5\}\}$ and consider a new simple game (N, v') such that

$$\mathcal{W}^{v'}_{\min} = \{\{1, 2\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 4, 5\}, \{2, 3, 4, 5\}\}.$$

Notice that the new simple game v' contains one more minimal winning coalition containing 2 but not 1, but the size of such a minimal winning coalition in v' is strictly larger than the size of any minimal winning coalition in v, and therefore is considered less likely to form. If a solution satisfying the property of independence of larger minimal winning coalitions ranks 1 strictly better than 2 in the simple game v, in v' the solution also must rank 1 strictly better than 2: the new (and larger) minimal winning coalition does not affect the strict ranking decided on the basis of smaller minimal winning coalitions.

In collective decision-making bodies, forming large winning coalitions in practice may result more difficult than forming small ones due to many factors, like the presence of complex institutional rules, the need of mediators in the decision-making process, higher negotiation costs or other "psychological" aspects, like contrasting political positions of their members. As a consequence, it is crucial to emphasize the impact of minimal winning coalitions of small size, as demanded by Property 3, which preserves strict rankings after the addition of large minimal winning coalitions.

Proposition 2 *Let R be a solution satisfying Properties 1 (DM) and 2 (AMWC). Then for any simple game v and $i, j \in N$ such that $\theta_v(i) = \theta_v(j)$ we have that $i I^v j$.*

Proof Since $\theta_v(i) = \theta_v(j)$, we have that $i_k = j_k$ for all $k \in \{1, \ldots, n\}$. Define a bijection π on $2^{N\setminus\{i,j\}}$ such that for each $k \in \{1, \ldots, n-1\}$ and for each coalition $S \in 2^{N\setminus\{i,j\}}$ of size $k-1$ with $S \cup \{j\} \in \mathcal{W}^v_{\min}$, $\pi(S) = T$, where $T \in 2^{N\setminus\{i,j\}}$ is a coalition of size $k-1$, with $T \cup \{i\} \in \mathcal{W}^v_{\min}$. Consider a game v_π such that $S \cup \{i\} \in \mathcal{W}^v_{\min} \Leftrightarrow S \cup \{i\} \in \mathcal{W}^{v_\pi}_{\min}$ and $S \cup \{j\} \in \mathcal{W}^v_{\min} \Leftrightarrow \pi(S) \cup \{j\} \in \mathcal{W}^{v_\pi}_{\min}$. So, we have that for all $k \in \{1, \ldots, n\}$ and all coalitions $T \in 2^{N\setminus\{i,j\}}$ of size $k-1$ with $T \cup \{i\} \in \mathcal{W}^v_{\min}$

$$T \cup \{i\} \in \mathcal{W}^{v_\pi}_{\min} \Leftrightarrow T \cup \{j\} \in \mathcal{W}^{v_\pi}_{\min}.$$

Then, $i \sim^{v_\pi} j$ (players i and j are equally desirable in v_π) and, by Property 1

$$i\, I^{v_\pi}\, j. \tag{3}$$

Notice that i, j, π, v and v_π satisfy the conditions for bijections demanded in the statement of Property 2, with v_π such that the set of minimal winning coalitions of v_π is

$$\mathcal{W}^{v_\pi}_{\min} = \bigcup_{T \in 2^{N\setminus\{i,j\}} \text{ s.t. } T\cup\{i\} \in \mathcal{W}^v_{\min}} \{T \cup \{i\}, T \cup \{j\}\}$$

(notice that the minimality of the elements in $\mathcal{W}^{v_\pi}_{\min}$ is guaranteed by the minimality of the elements in $\mathcal{W}^v_{\min}$). So, since R satisfies Property 2, we have that

$$i\, I^v\, j \Leftrightarrow i\, I^{v_\pi}\, j. \tag{4}$$

So, by relation (3), we have $i\, I^v\, j$, which concludes the proof. □

Theorem 1 *The LRM solution R_l is the unique solution that fulfils Properties 1 (DM), 2 (AMWC) and 3 (ILMWC).*

Proof By Proposition 1 we have that R_l fulfils Property 1 (DM). It is easy to check that it also fulfils Properties 2 (AMWC) and 3 (ILMWC) (it directly follows from Definition 1 and the lexicographic relation).

To show that R_l is the unique index fulfilling Properties 1 (DM), 2 (AMWC) and 3 (ILMWC), we need to prove that, if a solution $R : \mathcal{SG}^N \to \mathcal{T}^N$ satisfies Properties 1 (DM), 2 (AMWC) and 3 (ILMWC), then $i\, R^v_l\, j \Leftrightarrow i\, R^v\, j$ or, equivalently, $i\, P^v_l\, j \Leftrightarrow i\, P^v\, j$ and $i\, I^v_l\, j \Leftrightarrow i\, I^v\, j$.

We first prove that $i\, P^v_l\, j \Leftrightarrow i\, P^v\, j$:
($\Rightarrow$)
Let $i\, P^v_l\, j$. By Definition 3, let k' be the smallest integer in $\{1, \ldots, n\}$ with $i_{k'} > j_{k'}$. Let $s = i_{k'} - j_{k'}$ and $\mathcal{S}^i_{k'} = \{S \in \mathcal{W}^v_{\min} : |S| = k' \text{ and } S \cap \{i, j\} = i\}$ be a subset of coalitions in $\mathcal{W}^v_{\min}$ of size k' containing i but not j such that $|\mathcal{S}^i_{k'}| = s$. Moreover, let $\Sigma = \{S \in \mathcal{W}^v_{\min} : |S| > k'\}$ be the set of coalitions in $\mathcal{W}^v_{\min}$ with cardinality strictly larger than k'.

Consider a new simple game v' such that $\mathcal{W}^{v'}_{\min} = \mathcal{W}^v_{\min} \setminus \Sigma$, and the set of minimal winning coalitions containing j (of size at most k') in $\mathcal{W}^{v'}_{\min}$:

$$\mathcal{S}^j = \{S \cup \{j\} : S \in 2^{N\setminus\{i,j\}} \text{ with } S \cup \{j\} \in \mathcal{W}^{v'}_{\min}\}.$$

Define a bijection π on $2^{N\setminus\{i,j\}}$ such that for each $t \in \{1, \ldots, k'\}$ and for each coalition $S \in 2^{N\setminus\{i,j\}}$ of size $t-1$ with $S \cup \{j\} \in \mathcal{W}^{v'}_{\min}$, $\pi(S) = T$, where $T \in 2^{N\setminus\{i,j\}}$ is a coalition of size $t-1$, with $T \cup \{i\} \in \mathcal{W}^{v'}_{\min}$. So, the set of minimal winning coalitions contained in $\mathcal{S}^j$ after the transformation via π is:

$$\mathcal{T}^j = \{\pi(S) \cup \{j\} : S \in 2^{N\setminus\{i,j\}} \text{ with } S \cup \{j\} \in \mathcal{S}^j\}.$$

Consider a new game $\hat{v}_\pi$ such that

$$\mathcal{W}^{\hat{v}_\pi}_{\min} = \left(\mathcal{W}^{v'}_{\min} \setminus \mathcal{S}^j\right) \cup \mathcal{T}^j.$$

So, we have that for all coalitions $T \in 2^{N\setminus\{i,j\}}$ of size $t-1$, $t \in \{1, \ldots, k'\}$,

$$T \cup \{j\} \in \mathcal{W}^{\hat{v}_\pi}_{\min} \Rightarrow T \cup \{i\} \in \mathcal{W}^{\hat{v}_\pi}_{\min},$$

and, consequently, for all $S \in 2^{N\setminus\{i,j\}}$,

$$S \cup \{j\} \in \mathcal{W}^{\hat{v}_\pi} \Rightarrow S \cup \{i\} \in \mathcal{W}^{\hat{v}_\pi},$$

which means that $i \succeq^{\hat{v}_\pi} j$. So, by Property 1 (DM), we have that $i\, P^{\hat{v}_\pi}\, j$.

On the other hand,

$$S \cup \{i\} \in \mathcal{W}^{v'}_{\min} \Leftrightarrow S \cup \{i\} \in \mathcal{W}^{\hat{v}_\pi}_{\min}$$

and

$$S \cup \{j\} \in \mathcal{W}^{v'}_{\min} \Leftrightarrow \pi(S) \cup \{j\} \in \mathcal{W}^{\hat{v}_\pi}_{\min},$$

for all $S \in 2^{N\setminus\{i,j\}}$, and therefore, by Property 2 on R applied to v' and $\hat{v}_\pi$, we also have $i\, P^{v'}\, j$.

Finally, by Property 3 on R (with v' in the role of v in the statement of Property 3), we have that $i\, P^{v}\, j$, as $\mathcal{W}^{v}_{\min} = \mathcal{W}^{v'}_{\min} \cup \Sigma$.

($\Leftarrow$)

Let $i\, P^{v}\, j$. Suppose that $i\, I^{v}_{l}\, j$. Then, by Definition 3, $\theta_v(i) = \theta_v(j)$. So, by Proposition 2, $i\, I^{v}\, j$, which yields a contradiction with $i\, P^{v}\, j$. Since it can't even be $j\, P^{v}_{l}\, i$ (by the other implication proved above), and by the fact that P^{v}_{l} is a total relation, it must be $i\, P^{v}_{l}\, j$.

We now prove that $i\, I^{v}_{l}\, j \Leftrightarrow i\, I^{v}\, j$:

($\Rightarrow$)

Let $i\, I^{v}_{l}\, j$. Then, by Definition 3, $\theta_v(i) = \theta_v(j)$. So, by Proposition 2 and the fact that R^v satisfies Properties 1 and 2, $i\, I^{v}\, j$.

($\Leftarrow$)

Let $i\,I^v\,j$. As we have shown previously, $i\,P_l^v\,j \Leftrightarrow i\,P^v\,j$. So it is not possible that $i\,P_l^v\,j$ or $j\,P_l^v\,i$. Since P_l is a total relation, it must be $i\,I_l^v\,j$, which concludes the proof. □

We end this section showing the logical independence of Properties 1, 2 and 3.

Example 7 [No Property 1] Given $i, j \in N$, consider the ranking solution R_{DM} defined by

$$i\ R_{DM}^v\ j \qquad \text{iff} \qquad v(\{i\}) \geq v(\{j\}).$$

This solution satisfies all the Properties but Property 1.

Example 8 [No Property 2] For any $i \in N$, let $B(i)$ the the largest player index within minimal winning coalitions containing player i, i.e.

$$B^v(i) = \max_{S \in \mathcal{W}_{\min}^N : i \in S} \Big(\min_{j \in S \setminus \{i\}} j \Big).$$

Consider the ranking solution R_{AMWC}^v such that

$$\begin{cases} i\ I_{AMWC}^v\ j & \text{if } i \sim^v j, \\ i\ P_{AMWC}^v\ j & \text{if } \theta_v(i) >_L \theta_v(j), \\ i\ P_{AMWC}^v\ j & \text{if } (i,j) \notin \succeq^v, (j,i) \notin \succeq^v, \theta_v(i) = \theta_v(j) \text{ and } B^v(i) > B^v(j), \end{cases}$$

This solution satisfies all the Properties but Property 2.

[It is clear that R_{AMWC}^v satisfies properties 1 and 3. To see that R_{AMWC}^v does not satisfy Property 2, consider, for instance, games v and v_π of Example 5. As we noticed, a solution satisfying Property 2 should rank players 3 and 4 in the same way in both games v and v_π, However, since $B^v(3) = B^v(4) = 1$ in v and $B^{v_\pi}(3) = 1$ and $B^{v_\pi}(4) = 2$ (and the two players are not in a desirable relation in both games) we have that $3\ I_{AMWC}^v\ 4$, while $4\ P_{AMWC}^{v_\pi}\ 3$.]

Example 9 [No Property 3] For each $i \in N$, let $\overline{\theta}_v$ be the n-dimensional vector $\overline{\theta}_v(i) = (i_n, \ldots, i_1)$ associated to v. Given $i, j \in N$, consider the vector ranking solution R_{ILMWC}^v such that

$$\begin{cases} i\ P_{ILMWC}^v\ j & \text{if } i \succ^v j, \\ i\ I_{ILMWC}^v\ j & \text{if } \theta_v(i) = \theta_v(j), \\ i\ P_{ILMWC}^v\ j & \text{if } (i,j) \notin \succeq^v, (j,i) \notin \succeq^v \text{ and } \overline{\theta}_v(i) >_L \overline{\theta}_v(j), \end{cases}$$

This solution satisfies all the Properties but Property 3.

[It is easy to verify that R_{ILMWC}^v satisfies Properties 1 and 2. To see that R_{ILMWC}^v does not satisfy property 3, consider, for instance, games v and v' of Example 6. Notice that $1 \succ^v 2$ and, so, $1\ P_{ILMWC}^v\ 2$. However, in game v', $(1,2) \notin \succeq^{v'}$ and $(2,1) \notin \succeq^{v'}$ (1 and 2 are not in desirable relation), while

$$\overline{\theta}_{v'}(2) = (0, 1, 0, 1, 0) >_L (0, 0, 3, 1, 0) = \overline{\theta}_{v'}(1)$$

and therefore $2\ P^{v'}_{ILMWC}\ 1$.]

5 Duality

In this section we investigate the connections between the LRM solution and the criticality-based ranking introduced in Aleandri et al. (2021) to rank players in a simple game. In Aleandri et al. (2021) a ranking over players is defined according to the power of blocking the grand coalition to be winning. Given a simple game (N, v) a coalition $B \subseteq N$ is called *blocking* coalition for N if $v(N \setminus B) = 0$. Let $\mathcal{B}^v$ be the set of all blocking coalitions in the game (N, v) and let $\mathcal{B}^v_{\min}$ be the set of all minimal blocking coalitions $\mathcal{B}^v_{\min} = \text{Min}\, \mathcal{B}^v$. Denote by i^*_k the number of minimal blocking coalitions (for N) of size k containing player i, so $i^*_k = |\{B \in \mathcal{B}^v_{\min} : i \in B, |B| = k\}|$ for all $k \in \{1, \ldots, n\}$. For each $i \in N$, let $\theta^*_v(i)$ be the n-dimensional vector $\theta^*_v(i) = (i^*_1, \ldots, i^*_n)$ associated to v.

The criticality-based ranking is based on the idea that the smaller is the size of a blocking coalition, the larger is the influence on the blocking power of its members; the ranking of a player in terms of blocking power is positively correlated first to the size of minimal blocking coalitions the player belongs to and second to their number.

Definition 3 The *criticality-based solution* is the function $R_c : \mathcal{SG}^N \longrightarrow \mathcal{T}^N$ defined for any simple game $v \in \mathcal{SG}^N$ as

$$i\ R^v_c\ j \quad \text{if} \quad \theta^*_v(i)\ \geq_L\ \theta^*_v(j).$$

Let I^v_c and P^v_c be the symmetric part and the asymmetric part of R^v_c, respectively.

Example 10 Consider the simple game of Example 1 then we have that

$$\mathcal{B}^v_{\min} = \{\{1, 4\}, \{1, 2, 3\}, \{1, 3, 5\}, \{2, 3, 4\}, \{2, 3, 5\}\}.$$

Therefore,

$$\begin{aligned}
&\theta^*_v(1) = (0, 1, 2, 0, 0),\ \ \theta^*_v(2) = (0, 0, 3, 0, 0),\\
&\theta^*_v(3) = (0, 0, 4, 0, 0),\ \ \theta^*_v(4) = (0, 1, 1, 0, 0),\\
&\theta^*_v(5) = (0, 0, 2, 0, 0).
\end{aligned}$$

So, the criticality-based ranking is such that

$$1\ P^v_c\ 4\ P^v_c\ 3\ P^v_c\ 2\ P^v_c\ 5.$$

We first show that the LRM coincides with the criticality-based ranking of the dual game.

Proposition 3 *Let (N, v) be a simple game. Then $R_l^v = R_c^{v^*}$.*

Proof Given the simple game (N, v) its dual v^* is defined by

$$v^*(S) = v(N) - v(N \setminus S), \tag{5}$$

for each coalition $S \in 2^N$. The proposition follows recalling that $\mathcal{W}_{\min}^{v^*} = \mathcal{B}_{\min}^v$, as proved in Proposition 3 in Aleandri et al. (2021), and then $\theta_v = \theta_{v^*}^*$. □

On the other hand, it is also interesting to study under which conditions the LRM and the criticality-based ranking coincide. To this purpose, we analyse the behaviour of the desirability relation on a simple game v and its dual v^*.

Proposition 4 *Given a simple game (N, v) and the dual game (N, v^*) then, $\forall i, j \in N, i \neq j$*

$$i \succeq^v j \quad \Longleftrightarrow \quad i \succeq^{v^*} j.$$

Proof $\Rightarrow$ By hypothesis, for all $S \subseteq N \setminus \{i, j\}$, $S \cup \{j\} \in \mathcal{W}^v$ implies that $S \cup \{i\} \in \mathcal{W}^v$. We want to prove that for all $S \subseteq N \setminus \{i, j\}$ such that $N \setminus \{S \cup \{j\}\} \notin \mathcal{W}^v$ implies that $N \setminus \{S \cup \{i\}\} \notin \mathcal{W}^v$. Suppose that $N \setminus \{S \cup \{i\}\}$ is winning then define $T := N \setminus \{S \cup \{i, j\}\}$. We observe that $T \cup \{j\}$ is winning then $T \cup \{i\} = N \setminus \{S \cup \{j\}\}$ is winning, i.e. a contradiction.

$\Leftarrow$ By hypothesis, $\forall S \subseteq N \setminus \{i, j\}$, $N \setminus \{S \cup \{j\}\} \notin \mathcal{W}^v$ implies that $N \setminus \{S \cup \{i\}\} \notin \mathcal{W}^v$. We want to prove that for all $S \subseteq N \setminus \{i, j\}$, $S \cup \{j\} \in \mathcal{W}^v$ implies that $S \cup \{i\} \in \mathcal{W}^v$. Suppose $S \cup \{i\} \notin \mathcal{W}^v$ and let $T = N \setminus \{S \cup \{i, j\}\}$. We observe that on the one hand $N \setminus \{T \cup \{i\}\}$ is not winning, but on the other hand $N \setminus \{T \cup \{i\}\} = S \cup \{j\}$ is winning, i.e. a contradiction, and the proof is complete. □

Corollary 1 *Let (N, v) be a simple game such that the desirability relation is total. Then, $R_l^v = R_c^v$. Moreover the LRM and the criticality-based ranking are self-dual.*

Example 11 Consider the weighted majority game (N, v) of Example 4. The minimal winning and blocking coalitions are

$$\mathcal{W}_{\min}^v = \{\{1, 2\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 4, 5\}\},$$

$$\mathcal{B}_{\min}^v = \{\{1\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 4, 5\}\}.$$

We have that

$$\begin{aligned}
&\theta_v^*(1) = (1, 0, 0, 0, 0), \ \theta_v^*(2) = (0, 0, 3, 0, 0),\\
&\theta_v^*(3) = (0, 0, 2, 0, 0), \ \theta_v^*(4) = (0, 0, 2, 0, 0),\\
&\theta_v^*(5) = (0, 0, 2, 0, 0).
\end{aligned}$$

The criticality-based ranking is such that

$$1\ P_c^v\ 2\ P_c^v\ 3\ I_c^v\ 4\ I_c^v\ 5.$$

So, as expected for a weighted majority game in which the desirable relation is total, $R_l^v = R_c^v$.

With the purpose of ranking players in a simple game according to their influence in the process of forming blocking coalitions, it seems natural to look at a dual version of Property 1.

Property 4 *(Dual Desirable Monotonicity (DDM))* Let $i, j \in N$. For any $v \in \mathcal{SG}^N$, a solution R satisfies the *dual desirable monotonicity* property if

$$i \sim^{v^*} j \Rightarrow i I^v j,$$

and

$$i \succ^{v^*} j \Rightarrow i P^v j.$$

A solution satisfying Property 4 obeys to the desirability relation defined on dual games, specifying that i is at least as desirable as j if we can replace player j with player i in any blocking coalition (instead of in any winning one). In a similar fashion, Properties 2 and 3 can be reformulated as their following dual counterparts.

Property 5 *(Dual Anonymity of Minimal Winning Coalitions (DAMWC))* Let $i, j \in N$, $v, v_\pi \in \mathcal{SG}^N$ and let π be a bijection on $2^{N\setminus\{i,j\}}$ with $|\pi(S)| = |S|$ and such that

$$S \cup \{i\} \in \mathcal{W}_{\min}^{v^*} \Leftrightarrow S \cup \{i\} \in \mathcal{W}_{\min}^{v_\pi^*}$$

and

$$S \cup \{j\} \in \mathcal{W}_{\min}^{v^*} \Leftrightarrow \pi(S) \cup \{j\} \in \mathcal{W}_{\min}^{v_\pi^*},$$

for all $S \in 2^{N\setminus\{i,j\}}$. A solution R satisfies the *anonymity of minimal winning coalitions* property if

$$i R^v j \Leftrightarrow i R^{v_\pi} j.$$

Property 6 *(Independence of Larger Minimal Winning Coalitions in the Dual (ILMWCD))* Let $i, j \in N$. For any $v \in \mathcal{SG}^N$, let $h = \max\{|S| : S \in \mathcal{W}_{\min}^v$ and $S \cap \{i, j\} \neq \varnothing\}$ be the highest cardinality of coalitions in the set $\mathcal{W}_{\min}^{v^*}$ containing either i or j. Let $\mathcal{S}_h$ be a collection of (minimal) winning coalitions in the dual game v^* with cardinality strictly larger than h, i.e., $\mathcal{S}_h = \{S_1, \ldots, S_r\}$ such that $S_k \subseteq N$, $|S_k| > h$ for $k = 1, \ldots, r$ and there is no $Q \in \mathcal{W}_{\min}^{v^*} \cup \mathcal{S}_h$ with $Q \subset S_k$, for all $k \in \{1, \ldots, r\}$. A solution R satisfies the property of *independence of larger minimal winning coalitions in the dual* if

$$i P^v j \Rightarrow i P^{v'} j,$$

where v' is a simple game such that the set of minimal winning coalitions is obtained as $\mathcal{W}_{\min}^{v'^*} = \mathcal{W}_{\min}^{v^*} \cup \mathcal{S}_h$.

We can state the following result.

Theorem 2 *The solution $R_{l'}$ such that $R_{l'}^v = R_l^{v^*}$ for all $v \in \mathcal{W}_{\min}^v$ is the unique solution that fulfils Properties 4, 5 and 6.*

Proof The proof follows the same steps of the proof of Theorem 1, with v^* in the role of v. □

By Proposition 3, and the fact that $(v^*)^* = v$ (the dual of the dual of a game v equals game v), we have that $R_{l'} = R_c$. Moreover, by Proposition 4, we have that Properties 1 and 4 are equivalent. So, the following corollary holds.

Corollary 2 *The criticality-based solution R_c is the unique solution that fulfils Properties 1, 5 and 6.*

In Aleandri et al. (2021), the criticality-based solution has been axiomatically characterized using four properties, namely, *Players' Anonymity*, *Dual Coalitional Anonymity*, *Dual Monotonicity* and *Independence of Higher Cardinalities* (see Sects. 4 in Aleandri et al. (2021) for a formal definitions of these axioms). Notice that Property 5 coincides with the property of *Dual Coalitional Anonymity* in Aleandri et al. (2021), while Property 6 coincides with the property of Independence of Higher Cardinalities in Aleandri et al. (2021). So, according to Corollary 2, Property 1 replaces properties of Players' Anonymity and Dual Monotonicity in the axiomatic characterization of the criticality-based solution presented in Aleandri et al. (2021).

Example 12 Consider the simple game (N, v) in Example 2.7 in Alonso-Meijide and Freixas (2010):

$$\mathcal{W}_{\min}^v = \{\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,3,4\},\{3,4,5\}\};$$

$$\mathcal{B}_{\min}^v = \{\{1,3\},\{1,4\},\{1,5\},\{2,3\},\{2,4\},\{3,4,5\}\};$$

The LRM solution in v is

$$1\ P_\ell^v\ 2\ I_\ell^v\ 3\ I_\ell^v\ 4\ P_\ell^v\ 5.$$

and the LRM solution in v^* is

$$1\ P_\ell^{v^*}\ 3\ I_\ell^{v^*}\ 4\ P_\ell^{v^*}\ 2\ P_\ell^{v^*}\ 5.$$

Then if the desirability relation is not a total preorder the LRM solution in not self-dual. The same result holds for the criticality-based solution.

We conclude this section pointing out that the axioms of Dual Coalitional Monotonicity and of Players' Anonymity are replaced by Desirable Monotonicity (Property 1) in the characterization of the criticality-based solution in Aleandri et al. (2021), as

shown by Corollary 2, but the two axioms do not imply Property 1. In fact, as shown by Example 4, the ranking over players represented by the PGI does not satisfy Property 1 (we leave to the reader to check that such a ranking satisfy both Dual Coalitional Monotonicity and Players' Anonymity axioms).

6 Conclusions

In this paper, using the minimal winning coalitions of a simple game, we introduced a new ranking among players that satisfies the desirability relation: the Lexicographic Ranking based on Minimal winning coalitions. The players are ranked according to the size of the minimal winning coalitions they belong to and then to the number of such coalitions. The ranking solution satisfies the coalitional anonymity property and the independence of larger minimal winning coalitions property that together with a monotonicity property rooted on the desirability relation uniquely characterized it. Looking at the dual game, we prove that there is a relation between the Lexicographic Ranking based on Minimal winning coalitions and the criticality-based ranking and, consequently, between ranking players according to their power to win and to their power to block the grand coalition. In particular, if the desirable relation is total the two rankings coincide.

Following this line of research, it would be interesting to delve more into the connection between the power to initiate and the power to block a winning coalition Deegan and Packel (1980), in particular, when the desirability relation between two players does not hold.

Acknowledgements The authors want to thank professor Marco Dall'Aglio for the valuable discussions and the anonymous referee for the useful comments. S. Moretti gratefully acknowledges the support of the ANR project THEMIS (ANR-20-CE23-0018).

References

Aleandri, M., Dall'Aglio, M., Fragnelli, V., & Moretti, S. (2021). Minimal winning coalitions and orders of criticality. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04199-6

Alonso-Meijide, J. M., & Freixas, J. (2010). A new power index based on minimal winning coalitions without any surplus. *Decision Support Systems, 49*(1), 70–76.

Banzhaf, J. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19*, 317–343.

Carreras, F., & Freixas, J. (1995). Complete simple games. *Mathematical Social Sciences, 32*, 139–155.

Deegan, J., & Packel, E. W. (1980). An axiomated family of power indices for simple n-person games. *Public Choice, 35*, 229–239.

Deegan, J., & Packel, E. W. (1980). A new index of power for simple n-person games. *International Journal of Game Theory, 7*, 113–123.

Diffo Lambo, L., & Moulen, J. (2002). Ordinal equivalence of power notions in voting games. *Theory and Decision, 53*, 313–325.

Dubey, P., Neyman, A., & Weber, R. (1981). Value theory without efficiency. *Mathematics of Operations Research, 6*(1), 122–128.

Fertö, I., Kóczy, L. Á., Kovács, A., & Sziklai, B. R. (2020). The power ranking of the members of the Agricultural Committee of the European Parliament. *European Review of Agricultural Economics, 47*(5), 1897–1919.

Freixas, J., & Gambarelli, G. (1997). Common internal properties among power indices. *Control and Cybernetics, 26*, 591–604.

Freixas, J., Marciniak, D., & Pons, M. (2012). On the ordinal equivalence of the Johnston, Banzhaf and Shapley power indices. *European Journal of Operational Research, 216*(2), 367–375.

Freixas, J., & Pons, M. (2005). Two measures of circumstantial power: Influences and bribes. *Homo Oeconomicus, 22*, 569–588.

Freixas, J., & Pons, M. (2008). Circumstantial power: Optimal persuadable voters. *European Journal of Operational Research, 186*(3), 1114–1126.

Holler, M. J. (1982). Forming coalitions and measuring voting power. *Political Studies, 30*, 262–271.

Holler, M. J., & Packel, E. W. (1983). Power, luck and the right index. *Journal of Economics, 43*, 21–29.

Holler, M. J., & Nurmi, H. (Eds.). (2013). *Power, voting, and voting power: 30 years after*. Springer.

Isbell, J. R. (1958). A class of simple games. *Duke Mathematical Journal, 25*, 423–439.

Johnston, R. J. (1978). On the measurement of power: some reactions to Laver. *Environment and Planning A*, *l0*(8), 907–914.

Lindelauf, R. H., Hamers, H. J., & Husslage, B. G. M. (2013). Cooperative game theoretic centrality analysis of terrorist networks: The cases of jemaah islamiyah and al qaeda. *European Journal of Operational Research, 229*(1), 230–238.

Moretti, S., Patrone, F., & Bonassi, S. (2007). The class of microarray games and the relevance index for genes. *Top, 15*(2), 256–280.

Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal of Applied Mathematics, 17*, 1163–1170.

Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*, 787–792.

Taylor, A. D., & Zwicker, W. S. (1999). *Simple games: Desirability relations, trading, pseudoweightings*. Princeton University Press.

The Public Good Index for Games with Several Levels of Approval in the Input and Output

Sascha Kurz

1 Introduction

Assume that you are submitting a paper[1] to a computer science conference (or some other scientific discipline with a similar reviewing convention). Your paper is usually send to several reviewers, which are typically chosen by the programming committee or assign themselves in some kind of bidding procedure. Unattached the selection process, for each paper, there exists a set N of reviewers. The task of the reviewers is to read and to evaluate the submitted paper. Besides some comments and remarks in free text, a summarizing evaluation according to a certain predefined scale is requested. A typical scale consists examples of the possible answers "strong accept", "accept", "weak accept", "borderline", "weak reject", "reject", and "strong reject". After every reviewer has announced his or her evaluation, these individual opinions are summarized to a group decision, where we assume that only the outcomes "accept" or "reject" are possible. Of course, this oversimplifies the practical setting where we may have discussion rounds between the reviewers with the possibility to adjust their evaluations or some kind of interaction with the authors of the paper. Such a decision rule v may be formalized as follows: For some set of agents N and a set of levels of approval for the input J, each vector in $J^{|N|}$ is mapped to an element of the set of levels of approval in the output K. In our example, we have $|J| = 7$ and $|K| = 2$, but may also consider an output set K of cardinality three by distinguishing between a lecture, a poster presentation, or rejection. If the options in J can be mapped to a numerical score, e.g., $+3, +2, +1, 0, -1, -2, -3$ in our example, then such a decision rule might be simply given by some threshold τ, i.e., accept all papers with mean of the scores at least τ. However, rules might be more complicated including extra conditions, e.g., requiring that no paper with at least one

[1] The paper is dedicated to the occasion of the 75th birthday of Manfred J. Holler.

S. Kurz (✉)
University of Bayreuth, Bayreuth, Germany
e-mail: sascha.kurz@uni-bayreuth.de

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_14

"strong reject" is accepted. Given a specific decision rule v one might ask for the "influence" of a specific agent $i \in N$ on the group decision. Having only homogeneous agents in mind this question does not seem to make too much sense. However, agents may also be heterogeneous. In our example, the reviewers may have different levels of expertise, which is indeed a common query to the reviewer when writing his or her evaluation. Of course, we as the authors of the paper usually do not have the details to determine the influence of the individual reviewers and should have little interest to do so, but the author of the one day is the organizer of a huge conference the other day and possibly in charge to design the details of the decision rules.

Taking our exemplifying story aside, we can clearly imagine situations where the individual opinions of $|N|$ agents from an ordered set J of inputs are mapped to an output from an ordered set K. To this end, $(|J|, |K|)$ simple games have been introduced, see e.g., (Freixas and Zwicker, 2003, 2009), and we remark that simple games are in one-to-one correspondence to $(2, 2)$ simple games with $J = \{0, 1\}$ and $K = \{0, 1\}$. Measurements of influence for simple games are also called power indices and the Public Good index, introduced in Holler (1982), is a particular example. The question of this paper is whether a measure in the vein of the Public Good index can be defined for the class of (j, k) simple games. We motivate a few variants and give axiomatizations. An axiomatization of the Public Good index for simple games was given in Holler and Packel (1983), so that some people also speak of the Holler–Packel index, and the generalization to TU games was axiomatized in Holler and Li (1995). A different axiomatization, for both cases and based on potential functions, was given in Haradau and Napel (2007). For $(j, 2)$ simple games, a Public Good index was recently introduced in Courtin and Tchantcho (2020) along with two axiomatizations.

The remaining part of this paper is structured as follows. In Sect. 1.1, we summarize some necessary preliminaries from the literature before we discuss different generalizations of the Public Good index and corresponding axiomatizations to the class of (j, k) simple games in Sect. 1.2.

1.1 Preliminaries

Let $N = \{1, 2, ..., n\}$ be a finite set of agents or voters. Any subset S of N is called a coalition and the set of all coalitions of N is denoted by the power set 2^N. For given integers $j, k \geq 2$, we denote by $J = \{0, \ldots, j-1\}$ the possible input levels and by $K = \{0, \ldots, k-1\}$ the possible output levels, respectively. We write $x \leq y$ for $x, y \in \mathbb{R}^n$ if $x_i \leq y_i$ for all $1 \leq i \leq n$. For each $\emptyset \subseteq S \subseteq N$, we write x_S for the restriction of $x \in \mathbb{R}^n$ to $(x_i)_{i \in S}$. As an abbreviation, we write $x_{-S} = x_{N \setminus S}$. Instead of $x_{\{i\}}$ and $x_{-\{i\}}$, we write x_i and x_{-i}, respectively. Slightly abusing notation we write $\mathbf{a} \in \mathbb{R}^n$, for the vector that entirely consists of a's, e.g., $\mathbf{0}$ for the all-zero vector.

Definition 1 Let $j, k \geq 2$ and $n \geq 0$ be integers. A *(j,k) simple game* is a mapping $v\colon J^n \to K$ satisfying $v(\mathbf{0}) = 0$ and $v(x) \leq v(y)$ for all $x, y \in J^n$ with $x \leq y$.[2]

Example 1 For $n = j = k = 3$, let the (3, 3) simple game v be defined via

$$v(x) = \begin{cases} 0 : 3x_1 + 2x_2 + x_3 < 7 \\ 1 : 7 \leq 3x_1 + 2x_2 + x_3 < 12 \\ 2 : x_1 = x_2 = x_3 = 2 \end{cases}$$

for all $x \in \{0, 1, 2\}^3$.

Definition 2 A *simple game* is a mapping $v\colon 2^N \to \{0, 1\}$ that satisfies $v(\emptyset) = 0$, $v(N) = 1$, and $v(S) \leq v(T)$ for all $\emptyset \subseteq S \subseteq T \subseteq N$, where the finite set N is called the *player set* or *set of players*.[3]

Let v be a simple game with player set N. A subset $S \subseteq N$ is called *winning coalition* if $v(S) = 1$ and *losing coalition* otherwise. A winning coalition $S \subseteq N$ is called *minimal winning coalition* if all proper subsets $T \subsetneq S$ of S are losing. The set of minimal winning coalitions is denoted by $\mathrm{MWC}(v)$.

Example 2 For player set $N = \{1, 2, 3\}$, let v be the simple game defined by $v(S) = 1$ iff $w(S) := \sum_{i \in S} w_i \geq 3$ and $v(S) = 0$ otherwise for all $S \subseteq N$, where $w_1 = 3$, $w_2 = 2$, and $w_3 = 1$.

The winning coalitions of the simple game from Example 2 are given by $\{1\}$, $\{2, 3\}$, $\{1, 2\}$, $\{1, 3\}$, and $\{1, 2, 3\}$. Only $\{1\}$ and $\{2, 3\}$ are minimal winning coalitions.

In order to embed a given simple game $v\colon 2^N \to \{0, 1\}$ as a (2, 2) simple game $\hat{v}$ with $J = \{0, 1\}$ and $K = \{0, 1\}$, we assume $N = \{1, \ldots, n\}$. To each coalition $S \subseteq N$, we assign the vector $x^S \in \{0, 1\}^n$ with $x_i^S = 1$ iff $i \in S$ and $x_i^S = 0$ otherwise. Given a vector $x \in \{0, 1\}^n$ the corresponding coalition is given by $S = \{i \in N \mid x_i = 1\}$, so that $v(S) = \hat{v}(x^S)$.

The (raw) Public Good index for a simple game v with player set N and a player $i \in N$ is given by

$$\mathrm{PGI}_i(v) = |\{S \in \mathrm{MWC}(v) \mid i \in S\}|. \tag{1}$$

With this, the (normalized) Public Good index is given by

$$\overline{\mathrm{PGI}_i}(v) = \frac{\mathrm{PGI}_i(v)}{\sum_{j \in N} \mathrm{PGI}_j(v)} \tag{2}$$

[2] Some authors also require $v(\mathbf{j} - \mathbf{1}) = k - 1$, which would clash with the potential function approach as it is the case for simple games. Note that we have reversed the order of the input levels of approval compared to Freixas and Zwicker (2003).

[3] In some papers $v(S) \leq v(T)$ is dropped in the definition of a simple game and they speak of *monotonic simple games* is additionally assumed. For the potential function approach, we will drop the condition $v(N) = 1$ later on, while it is indeed necessary for the normalized Public Good index.

and is, e.g., *efficient*, i.e., $\sum_{i\in N}\overline{\mathrm{PGI}_i}(v)=1$. Note that for the normalized version it is important to assume that $v(N)=1$ since $\mathrm{MWC}(v)$ is empty otherwise, so that $\overline{\mathrm{PGI}_i}(v)$ would be undefined.

A generalization of simple games, without the monotonicity assumption, are games with transferable utility—so-called TU games.

Definition 3 A *TU game* is a mapping $v\colon 2^N\to\mathbb{R}$ with $v(\emptyset)=0$, where the finite set N is called the *player set* or *set of players*.

If we additionally assume $v(S)\leq v(T)$ for all $\emptyset\subseteq S\subseteq T\subseteq N$, we speak of a *monotone TU game* or a *capacity*.

The analog of minimal winning coalitions in the context of TU games are *minimal crucial coalitions*, see e.g. (Haradau and Napel, 2007) or *real gaining coalitions*, see (Holler and Li, 1995). To this end, we call a player $i\in S\subseteq N$ *crucial* in a TU game v if $v(S)>v(S\backslash i)$. A coalition S in which every player i is crucial is called *minimal crucial coalition* and the set of minimal crucial coalitions is denoted by $\mathrm{MCC}(v)$. A coalition $S\subseteq N$ is called a real gaining coalition if $v(S)-v(T)>0$ for all proper subsets $\emptyset\subseteq T\subsetneq S$ of S. The set of all real gaining coalitions of v is denoted by $\mathrm{RGC}(v)$. Note that for monotone TU games there is no difference between a minimal crucial and a real gaining coalition, i.e., $\mathrm{MCC}(v)=\mathrm{RGC}(v)$. With these generalized notions, the Public Good value for a TU game v with player set N and a player $i\in N$ is given by

$$\mathrm{PGV}_i(v)=\sum_{S\in\mathrm{MCC}(v),i\in S}v(S), \tag{3}$$

so that $\mathrm{PGI}_i(v)=\mathrm{PGV}_i(v)$ if v is a simple game. For the rest[4] of the article, we will refer to $\mathrm{MCC}(v)$ as the minimal critical coalitions of v.

Let Γ be a subclass of all TU games. A *value* on Γ is a function Ψ that maps each game $v\in\Gamma$ to $\mathbb{R}^{|N|}$, where N is the player set of v. An example of a value is the Public Good value PGV, defined componentwise in Equation (3). A *potential* on Γ is a function P that maps each game $v\in\Gamma$ to a real number $P(v)$.

Definition 4 A value Ψ on Γ *admits a potential function* if there exists a potential $P\colon\Gamma\to\mathbb{R}$ such that

$$\Psi_i(v)=P(v)-P(v_{-i}) \tag{4}$$

for all $v\in\Gamma$ and all $i\in N$, where N is the player set of v and v_{-i} is the TU game with player set $N\backslash\{i\}$ defined by $v_{-i}(S)=v(S)$ for all $\emptyset\subseteq S\subseteq N\backslash\{i\}$.

[4] Note that the authors from Holler and Li (1995) used the definition $\mathrm{PGV}_i(v)=\sum_{S\in\mathrm{RGC}(v),i\in S}v(S)$, while the authors from Haradau and Napel (2007) used $\mathrm{PGV}_i(v)=\sum_{S\in\mathrm{MCC}(v),i\in S}v(S)$. As already mentioned, there is no difference for monotone TU games. Also, the axiomatization of the Public Good value from Haradau and Napel (2007) can be slightly adjusted by replacing the notion of minimal critical coalitions by real gaining coalitions in their definition of $\pi(v,N)$ and the corresponding axiom of *distributing the worths* of MCCs.

Note that the subclass Γ of TU games has to be closed with respect to taking subgames v_{-i} in order to apply this definition. So, from a technical point of view, we either have to include the game $v_\emptyset$ with an empty player set in the set of TU games and subclasses of TU games Γ or define $P(v_\emptyset) := 0$ separately (which is the usual choice).[5] As shown in (Haradau and Napel, 2007, Proposition 1), the Public Good value PGV admits a potential P on the class Γ of (monotone) TU games, where

$$P(v) = \sum_{S \in \mathrm{MCC}(v)} v(S). \tag{5}$$

Note that each minimal critical coalition S in v with $i \notin S$ is also a minimal critical coalition in v_{-i} and vice versa. Analogously, each real gaining coalition S in v with $i \notin S$ is also a real gaining coalition in v_{-i} and vice versa.

We say that a value Ψ on Γ *distributes the sum of the worths of the minimal critical coalitions for all players* in v iff

$$\sum_{i \in N} \Psi_i(v) = \sum_{i \in N} \sum_{S \in \mathrm{MCC}(v), i \in S} v(S) = \sum_{S \in \mathrm{MCC}(v)} |S| \cdot v(S) \tag{6}$$

for all $v \in \Gamma$, where N is the player set of v. With this, (Haradau and Napel, 2007, Proposition 2) states that the Public Good value PGV is the unique value that admits a potential and distributes the sum of the worths of the minimal critical coalitions for all players on the class of monotone TU games. The great advantage of an axiomatization via a potential is that this also gives an axiomatization for all subclasses Γ' of TU games that are closed with respect to taking subgames v_{-i}. So, if we relax the condition $v(N) = 1$ of a simple game, we also obtain an axiomatization for simple games. Note that, while $v(N) = 1$, it may happen that $v_{-i}(N \backslash \{i\}) \neq 1$, i.e., v_{-i} does not contain a winning coalition, which happens if player i is a so-called *vetoer*.

Another common property of values is linearity. To this end, we note that TU games form an $\mathbb{R}$-vector space with sum $(v + v')(S) := v(S) + v'(S)$ and scalar multiplication $(\lambda \cdot v)(S) := \lambda \cdot v(S)$ for all TU games v, v' with the same player set N, all $\lambda \in \mathbb{R}$, and all $S \subseteq N$. With this, a value Ψ is called *linear* if $\Psi(v + v') = \Psi(v) + \Psi(v')$ and $\Psi(\lambda \cdot v) = \lambda \cdot \Psi(v)$. From Equation (3), we can directly conclude that the Public Good value PGV is linear. If only the first property, on the sum of two TU games holds, then one speaks of *additivity*. Since the sum of two simple game (considered as TU games) does not need to be a simple game, the so-called *transfer axiom* was introduced by Dubey (1975):

$$\Psi(v \wedge v') + \Psi(v \vee v') = \Psi(v) + \Psi(v'),$$

where $\left(v \wedge v'\right)(S) = \min\{v(S), v'(S)\}$ and $\left(v \vee v'\right)(S) = \max\{v(S), v'(S)\}$ for all simple games v, v' with the same player set N and all coalitions $S \subseteq N$. Note that

[5] If we do not set $P(v_\emptyset) = 0$, then the potential of a value is only determined up to an additive constant.

the definition of $\wedge$ and $\vee$ might also be applied to general TU games. In our context, we only use $v \oplus v' := v \vee v'$ for two simple or TU games v, v'. Two simple games v and v' are called *mergeable* if $S \in \mathrm{MWC}(v)$ and $S' \in \mathrm{MWC}(v')$ implies $S \not\subseteq S'$ and $S' \not\subseteq S$. The identity $\mathrm{PGI}_i(v \oplus v') = \mathrm{PGI}_i(v) + \mathrm{PGI}_i(v')$ for the raw Public Good index for two mergeable simple games was used in Holler and Packel (1983) to axiomatize the normalized Public Good index. Similarly, for two (j, k) games v and v', we define $(v \oplus v')(x) = \max\{v(x), v(x')\}$ for all $x \in J^n$, where n is the number of players of v and v'.

1.2 *Generalizing the Public Good Index to* (j, k) *Simple Games*

The first question we have to answer is that for a suitable generalization of the concept of a minimal winning coalition in a simple game to an arbitrary (j, k) simple game. Having the definition of minimal critical and real gaining coalitions for TU games in mind, we propose the following:

Definition 5 Let v be a (j, k) simple game with player set $N = \{1, \ldots, n\}$ and $J = \{0, 1, \ldots, j-1\}$. A vector $x \in J^n$ is called *minimal critical* if $v(x) > v(x')$ for all $x' \in J^n$ with $x' \leq x$ and $x' \neq x$. The set of minimal critical vectors of v is denoted by $\mathrm{MCV}(v)$.

Note that for $j = 2$ and $k = 2$ each minimal critical vector x corresponds to a minimal winning coalition $S = \{1 \leq i \leq n \mid x_i = 1\}$ in the corresponding simple game. For $j = 2$ and arbitrary $k \geq 2$, we can embed a $(2, k)$ simple game v as a TU game $\hat{v}$, so that the minimal critical vectors of v are in 1-to-1 correspondence with the minimal critical coalitions of $\hat{v}$.

Let Γ be a subclass of all (j, k) simple games, where $j \geq 2$ and $k \geq 2$ are arbitrary but fixed. A *value* on Γ is a function Ψ that maps each game $v \in \Gamma$ to $\mathbb{R}^{|N|}$, where N is the player set of v. A *potential* on Γ is a function P that maps each game $v \in \Gamma$ to $\mathbb{R}$.

Definition 6 A value Ψ on a subclass Γ of (j, k) simple games *admits a potential function* if there exists a potential $P : \Gamma \to \mathbb{R}$ such that

$$\Psi_i(v) = P(v) - P(v_{-i}) \tag{7}$$

for all $v \in \Gamma$ and all $i \in N$, where N is the player set of v and v_{-i} is the (j, k) simple game with player set $N \backslash \{i\}$ defined by $v_{-i}(x) = v(y)$ for all $x \in J^{N \backslash \{i\}}$ and $y \in J^N$ with $y_i = 0$ and $y_j = x_j$ for all $j \in N \backslash \{i\}$.[6] Moreover, we set $P(v_\emptyset) := 0$ for a game $v_\emptyset$ with empty player set.

[6] By A^B, we denote the set of all mappings from B to A whose cardinality is $|A|^{|B|}$.

Again, the subclass Γ of (j,k) simple games has to be closed with respect to taking subgames v_{-i} in order to apply this definition. We observe that each minimal critical vector x of v with $x_i = 0$ is also a minimal critical vector of v_{-i} if we remove the entry for x_i (so that it is a vector in $J^{N\setminus\{i\}}$) and vice versa. We say that a value Ψ on a subclass Γ of (j,k) simple games *distributes the sum of the worths of the minimal critical vectors for all players* in v iff

$$\sum_{i=1}^{n}\Psi_i(v) = \sum_{i=1}^{n}\sum_{x\in \mathrm{MCV}(v),x_i\neq 0} v(x) = \sum_{x\in \mathrm{MCV}(v)} v(x)\cdot\left|\{1\le i\le n \mid x_i\neq 0\}\right| =: \Lambda(v) \tag{8}$$

for all $v\in\Gamma$, where $N=\{1,\ldots,n\}$ is the player set of v.

Theorem 1 *Let $j,k\ge 2$ be integers. Then, there exists a unique value Ψ on the class Γ of all (j,k) simple games that admits a potential function and distributes the sum of the worths of the minimal critical vectors for all players. We have*

$$\Psi_i(v) = \sum_{x\in \mathrm{MCV}(v),x_i\neq 0} v(x) \tag{9}$$

for all $v\in\Gamma$ and all i in the player set $\{1,\ldots,n\}$ of v. The potential function is given by

$$P(v) = \sum_{x\in \mathrm{MCV}(v)} v(x) \tag{10}$$

for all $v\in\Gamma$.

Proof First, we assume that the potential is given by Eq. (10). Since Ψ admits a potential function, we have

$$\begin{aligned}\Psi_i(v) = P(v) - P(v_{-i}) &= \sum_{x\in \mathrm{MCV}(v)} v(x) - \sum_{x\in \mathrm{MCV}(v_{-i})} v_{-i}(x)\\ &= \sum_{x\in \mathrm{MCV}(v)} v(x) - \sum_{x\in \mathrm{MCV}(v_{-i})} v(x)\\ &= \sum_{x\in \mathrm{MCV}(v),x_i\neq 0} v(x)\end{aligned}$$

for all $v\in\Gamma$ and all i in the player set of v, where we have used the relation between the minimal critical vectors of v and those of v_{-i}. Thus, Eq. (9) is valid. With this, we have

$$\sum_{i=1}^{n}\Psi_i(v) = \sum_{i=1}^{n}\sum_{x\in \mathrm{MCV}(v),x_i\neq 0} v(x) = \sum_{x\in \mathrm{MCV}(v)} v(x)\cdot\left|\{1\le i\le n \mid x_i\neq 0\}\right| = \Lambda(v),$$

i.e., Ψ distributes the sum of the worths of the minimal critical vectors for all players and so satisfies both axioms.

For the other direction, we assume that Ψ admits a potential $\tilde{P}$ so that

$$\Lambda(v) = \sum_{i=1}^{n} \Psi_i(v) = \sum_{i=1}^{n} \left(\tilde{P}(v) - \tilde{P}(v_{-i})\right) = n \cdot \tilde{P}(v) - \sum_{i=1}^{n} \tilde{P}(v_{-i}),$$

which is equivalent to

$$\tilde{P}(v) = \frac{\Lambda(v) + \sum_{i=1}^{n} \tilde{P}(v_{-i})}{n} \tag{11}$$

for each $v \in \Gamma$, where $N = \{1, \dots, n\}$ is the player set of v. For each $S \subseteq N$, we denote by v_S the (j,k) simple game with player set S defined by $v_S(x) = v(y)$ for all $x \in J^S$, where $y \in J^N$ with $y_j = x_j$ for all $j \in S$ and $y_j = 0$ otherwise. For example, $v_{-1} = v_{N\setminus\{i\}}$ and $v_N = v$. Since $(v_S)_T = v_T$, for all $\emptyset \subseteq T \subseteq S \subseteq N$ Eq. (11) can be generalized to

$$\tilde{P}(v_S) = \frac{\Lambda(v_S) + \sum_{i \in S} \tilde{P}(v_{S\setminus\{i\}})}{|S|}$$

for all $\{i\} \subseteq S \subseteq N$. So, starting from $\tilde{P}(v_\emptyset) = 0$, we can recursively compute $\tilde{P}(v_S)$ for all $\emptyset \neq S \subseteq N$, so that especially $\tilde{P}(v) = \tilde{P}(v_N)$ is uniquely defined. □

We call the value Ψ for (j,k) simple games defined by Eq. (9) *Public Good value* (for (j,k) simple games). For the (3, 3) simple game v from Example 1 the minimal critical vectors are (1, 1, 2), (1, 2, 0), (2, 0, 1), (2, 1, 0), and (2, 2, 2), where $v(x) = 1$, for all $x \in \mathrm{MCV}(v)\setminus\{(2,2,2)\}$ and $v((2,2,2)) = 2$. With this, we compute

$$\Psi_1(v) = 6, \quad \Psi_2(v) = 5, \quad \text{and } \Psi_3(v) = 4$$

for the value Ψ characterized in Theorem 1.

We would like to remark that we also may motivate a different definition for a Public Good value for (j,k) simple games. To this end, we define the vector $y = x \downarrow i \in J^n$ for each $x \in J^n$ with $x_i \neq 0$ by $y_j = x_j$ for all $j \neq i$ and $y_i = x_i - 1$. Assume that agent i has strictly increasing costs in i and that the rewards are strictly increasing in $v(x)$.[7] As in the process of a coalition forming member by member, we may imagine that, starting from $x = \mathbf{0}$, the final vector x forms step by step via the inverse operation of $\downarrow$.[8] So, similarly, as one can argue that only minimal winning coalitions will be formed, we deduce that, under the described model for every finally formed vector $x \in J^n$ with $v(x) \neq 0$, we have $x \in \mathrm{MCV}(v)$. Now, what is the contribution of a player i to a minimal critical vector x with $x_i \neq 0$ to the worth $v(x)$? If the answer is $v(x)$, then we end up with the value characterized in Theorem 1. However, if we have a look at the minimal critical vector $x = (2,2,2)$ in the (3, 3)

[7] For (2, 2) simple games represented as simple games this means that entering a coalition comes at a certain cost while a coalition gets a reward iff it is a winning coalition.

[8] More precisely, for each $x \in J^n$ with $x_i \neq j - 1$, we can define the vector $y = x \uparrow i \in J^n$ by $y_j = x_j$ for all $j \neq i$ and $y_i = x_i + 1$.

simple game v from Example 1, then $v(1,2,2) = v(2,1,2) = v(2,2,1) = 1$ may justify the assumption that every player contributes just a surplus of 1 to the worth of vector x. Thus, we would obtain a value defined by

$$\Psi_i(v) = \sum_{x \in \mathrm{MCV}(v), x_i \neq 0} \big(v(x) - v(x \downarrow i)\big). \tag{12}$$

Note the similarity to the Banzhaf index. For simple games, the difference is that we sum over all minimal winning instead of all winning coalitions. For the (3, 3) simple game v from Example 1, we would obtain

$$\Psi_1(v) = 5, \quad \Psi_2(v) = 4, \quad \text{and } \Psi_3(v) = 3.$$

We observe that there is no difference between both variants if $k = 2$. And indeed, they match the variant introduced in Courtin and Tchantcho (2020). For all (j,k) simple games not identically mapping to zero, we define the normalized version

$$\overline{\Psi}_i(v) = \frac{\Psi_i(v)}{\sum_{j=1^n} \Psi_j(v)}. \tag{13}$$

Excluding the (j,k) simple games $v \equiv 0$, we speak of *non-trivial* (j,k) simple games. Our next aim is an axiomatization for $\overline{\Psi}$. To this end, we propose a generalization of mergeability for simple games:

Definition 7 Two (j,k) simple games v and v' with the same player set $\{1,\ldots,n\}$ are *mergeable* if

(1) $\mathrm{MCV}(v) \cap \mathrm{MCV}(v') = \emptyset$;
(2) $x \in \mathrm{MCV}(v), x' \in \mathrm{MCV}(v'), x \leq x' \quad \Rightarrow \quad v(x) < v'(x')$; and
(3) $x \in \mathrm{MCV}(v), x' \in \mathrm{MCV}(v'), x \geq x' \quad \Rightarrow \quad v(x) > v'(x')$.

Note that (2) and (3) imply (1). Since $v(x) > 0$, for all $x \in \mathrm{MCV}(v)$, the definition for (2, 2) simple games goes in line with the definition for simple games. Actually, we have $v(x) = 1$ for every minimal critical vector of some $(j,2)$ simple game. If $k > 2$, then we have to distinguish the critical vectors according to their output value $v(x)$. Next, we study the relation of the minimal critical vectors of the sum of two mergeable (j,k) simple games with those of their "summand games".

Lemma 1 *Let v be a (j,k) simple game with player set $\{1,\ldots,n\}$. For each vector $x \in J^n$ with $v(x) > 0$, there exists a vector $x' \leq x$ with $v(x') = v(x)$ and $x' \in \mathrm{MCV}(v)$.*

Proof If $x \in \mathrm{MCV}(v)$, then the statement is true for $x' = x$. Otherwise, there exists a player $1 \leq i \leq n$ with $x_i \neq 0$ such that $v(x) = v(x \downarrow i)$. If $x \downarrow i \in \mathrm{MCV}(v)$, then we can set $x' = x \downarrow i$ and are done. Otherwise, we iteratively apply the operator $\downarrow$ (which terminates since the number of players and output levels is finite). □

We remark that the minimal critical vector x' does not need to be unique. To this end, we may slightly adjust the $(3,3)$ simple game v from Example 1 by setting $v(x) = 1$ for $x = (2,2,2)$.

Lemma 2 *Let v and v' be two (j,k) simple games with the same player set $\{1,\dots,n\}$ that are mergeable. Then, we have*

$$\mathrm{MCV}(v \oplus v') = \mathrm{MCV}(v) \cup \mathrm{MCV}(v').$$

Proof Consider $x \in \mathrm{v} \oplus \mathrm{v}'$. Since $(v \oplus v')(x) = \max\{v(x), v'(x)\}$, we assume $(v \oplus v')(x) = v(x)$ and $v'(x) \le v(x)$ w.l.o.g. If $x \notin \mathrm{MCV}(v)$, then there exists a player $1 \le i \le n$ with $x_i \neq 0$ such that $v(x \downarrow i) = v(x)$. However, this implies $(v \oplus v')(x \downarrow i) \ge v(x \downarrow i) = v(x) = (v \oplus v')(x)$, which is a contradiction. Thus, we have $\mathrm{MCV}(v \oplus v') \subseteq \mathrm{MCV}(v) \cup \mathrm{MCV}(v')$.

Consider $x \in \mathrm{MCV}(v)$. First, we show $v(x) > v'(x)$. To this end, we apply Lemma 1 to conclude the existence of a vector $x' \in J^n$ with $x' \le x$ and $v'(x') = v'(x)$. Now the stated inequality is implied by Definition 7.(3) and we have $(v \oplus v')(x) = v(x)$. Assume $x \notin \mathrm{MCV}(v \oplus v')$ for a moment. Let $1 \le i \le n$ be a player with $(v \oplus v')(x \downarrow i) = (v \oplus v')(x)$. Since

$$(v \oplus v')(x \downarrow i) = \max\{v(x \downarrow i), v'(x \downarrow i)\} \le \max\{v(x \downarrow i), v'(x)\} < v(x) = (v \oplus v')(x),$$

we obtain a contradiction. Thus, $\mathrm{MCV}(v) \subseteq \mathrm{MCV}(v \oplus v')$ and, by symmetry, also $\mathrm{MCV}(v') \subseteq \mathrm{MCV}(v \oplus v')$, so that $\mathrm{MCV}(v) \cup \mathrm{MCV}(v') \subseteq \mathrm{MCV}(v \oplus v')$. □

Note that $\mathrm{MCV}(v) \cap \mathrm{MCV}(v') = \emptyset$, i.e., we have the disjoint union $\mathrm{MCV}(v \oplus v') = \mathrm{MCV}(v) \uplus \mathrm{MCV}(v')$.

We say that a minimal critical vector $x \in \mathrm{MCV}(v)$ is *critical for player i and output level τ* if $v(x) \ge \tau$ and $v(x \downarrow i) < \tau$. So, a given minimal critical vector $x \in \mathrm{MCV}(v)$ (with $x_i \neq 0$) is critical for $v(x) - v(x \downarrow i)$ output levels. Denoting the number of pairs (x, τ) such that $x \in \mathrm{MCV}(v)$ with $x_i \neq 0$ is critical for player i with output level τ by $c_i(v)$, we have

$$c_i(v \oplus v) = c_i(v) + c_i(v') \tag{14}$$

for two mergeable (j,k) simple games v, v' with player set $\{1,\dots,n\}$ and $1 \le i \le n$.

Definition 8 Let v be a (j,k) simple game with player set $\{1,\dots,n\}$. A player $1 \le i \le n$ is called a *null player* if we have $v(x) = v(x')$ for all $x, x' \in J^n$ with $x_j = x'_j$ for all $j \neq i$.

Note that we have $x_i = 0$ for every null player i and every minimal critical vector $x \in \mathrm{MCV}(v)$. The analog for simple games is that no null player is part of a minimal winning coalition.

Definition 9 Let v be a (j,k) simple game with player set $N := \{1,\dots,n\}$ and $\pi : N \to N$ be a permutation, i.e., a bijection. The (j,k) simple game πv is defined by $(\pi v)(x) = v(x')$ for all $x \in J^n$, where $x'_i = x_{\pi(i)}$ for all $1 \le i \le n$.

A value Φ on the class of (non-trivial) (j,k) simple games is called *anonymous* if for each permutation $\pi : N \to N$, we have $\overline{\Psi}_i(\pi v) = \overline{\Psi}_{\pi(i)}(v)$, where N is the player set of an arbitrary (non-trivial) (j,k) simple game v and $i \in N$ an arbitrary player.

Theorem 2 *The value $\overline{\Psi}$ defined in Eqs. (13) and (12) is the unique value for non-trivial (j,k) simple games that satisfies the axioms:*

(A1) *i is a null player in $v \;\Rightarrow\; \overline{\Psi}_i(v) = 0$.*
(A2) *$\overline{\Psi}$ is efficient, i.e., $\sum_{i=1}^{n} \overline{\Psi}_i(v) = 1$.*
(A3) *If $\mathrm{MCV}(v) = \{x\}$ for a game v, then $\overline{\Psi}_i(v) = \overline{\Psi}_j(v)$ for all players i, j with $x_i, x_j \neq 0$.*
(A4) *For all mergeable (j,k) simple games v, v' with player set N, we have*

$$\overline{\Psi}_i(v \oplus v') = \frac{c(v) \cdot \overline{\Psi}_i(v) + c(v') \cdot \overline{\Psi}_i(v')}{c(v) + c(v')}$$

for all $i \in N$, where $c(\tilde{v}) = \sum_{j \in N} c_j(\tilde{v})$ for every non-trivial (j,k) simple game $\tilde{v}$ with player set N.

Proof It is immediate that the value $\overline{\Psi}$ defined in Eqs. (13) and (12) satisfies the axioms (A1), (A2), and (A3). For (A4), we first note $\Psi_i(\tilde{v}) = c_i(\tilde{v})$ for every (j,k) simple game $\tilde{v}$ and every player i in $\tilde{v}$. Using the mergeability of v and v', we compute

$$\overline{\Psi}_i(v \oplus v') = \frac{c_i(v \oplus v')}{c(v \oplus v')} = \frac{c_i(v) + c_i(v')}{c(v) + c(v')} = \frac{c(v) \cdot \overline{\Psi}_i(v) + c(v') \cdot \overline{\Psi}_i(v')}{c(v) + c(v')}.$$

Conversely, given any value Φ on the class of non-trivial (j,k) simple games satisfying the axioms (A1) through (A4), we proceed as follows. First, we consider an arbitrary non-trivial (j,k) simple game v with $|\mathrm{MCV}(v)| = 1$ and let x be the unique minimal critical vector. From (A1), (A2), and (A3), we conclude

$$\Phi_i(v) = \begin{cases} 1 \,/\, |\{j \mid x_j \neq 0\}| & \text{if } x_i \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Now consider any non-trivial (j,k) simple game $\tilde{v}$ with player set N and minimal critical vectors enumerated as $\mathrm{MCV}(\tilde{v}) = \{x^1, \ldots, x^m\}$. Denoting the non-trivial (j,k) simple game with unique minimal critical vector x^h by v^h, where $1 \leq h \leq m$, we can write

$$\tilde{v} = v^1 \oplus v^2 \oplus \cdots \oplus v^m.$$

Note that the v^h are sequentially mergeable in the sense that v^{h+1} and $v^1 \oplus \cdots \oplus v^h$ are mergeable for each $h = 1, 2, \ldots, m-1$. We can extend (A4) inductively to a sum of such games to obtain for each player $i \in N$

$$\Phi_i(\tilde{v}) = \sum_{h=1}^{m} c(v^h)\Phi_i(v^h) / \sum_{h=1}^{m} c(v^h).$$

Thus, the axioms (A1)–(A4) allow us to compute $\Phi_i(\tilde{v})$ for each non-trivial (j,k) simple game $\tilde{v}$ and each player i of $\tilde{v}$, i.e., there is at most one value satisfying axioms (A1)–(A4). So, given our first observation on $\overline{\Psi}$, we conclude $\Phi = \overline{\Psi}$. □

We remark that the axioms (A1) and (A2) mimic similar axioms for simple or TU games that are used frequently in the literature. For axiom (A4), we refer to the discussion in Holler and Packel (1983) noting that the proof of Theorem 2 is rather similar to the one of (Holler and Packel, 1983, Sect. III). Note that, for $k = 2$ output levels, axiom (A3) can be replaced by anonymity, see Definition 9. However, for $k > 2$, we need some kind of stronger axiom in order to uniquely define the value of non-trivial (j,k) simple games with a unique minimal critical vector. Of course, axiom (A3) might be considered to be too demanding for the cases where $x_i, x_j \neq 0$ and $x_i \neq x_j$. There is an ongoing discussion about properties that a reasonable power index or value should have, see e.g. (Allingham, 1975; Kurz, 2020). We would also like to point the reader to the two axiomatizations of the Public Good index for $(j,2)$ simple games in Courtin and Tchantcho (2020), which share several axioms on the one hand and use a few different on the other hand.

Another approach to motivate the definition of a value for (j,k) simple games is pursued in Kurz et al. (2021) for the Shapley value.

Definition 10 Let v be an arbitrary (j,k) simple game with player set $N = \{1, \ldots, n\}$. The *average game*, denoted by $\widetilde{v}$, associated to v is defined by

$$\widetilde{v}(S) = \frac{1}{j^n(k-1)} \sum_{x \in J^n} \left[v((\mathbf{j}-\mathbf{1})_S, x_{-S}) - v(\mathbf{0}_S, x_{-S}) \right] \tag{15}$$

for all $S \subseteq N$.

For the $(3,3)$ simple game v from Example 1, the average game $\tilde{v}$ is given by $\tilde{v}(\emptyset) = 0$, $\tilde{v}(\{1\}) = \frac{1}{2}$, $\tilde{v}(\{2\}) = \frac{5}{18}$, $\tilde{v}(\{3\}) = \frac{1}{6}$, $\tilde{v}(\{1,2\}) = \frac{2}{3}$, $\tilde{v}(\{1,3\}) = \frac{2}{3}$, $\tilde{v}(\{2,3\}) = \frac{1}{2}$, and $\tilde{v}(\{1,2,3\}) = 1$. Note that $\tilde{v}$ always is a TU game taking values between 0 and 1.

In (Kurz et al., 2021, Theorem 4.1), it was shown that the Shapley value of a (j,k) simple game v, as defined in, e.g., (Freixas, 2005), equals the Shapley value of the TU game $\tilde{v}$. Unfortunately, there is no such nice relation between the Public Good value and our analogs for (j,k) simple games since for the $(3,3)$ simple game from Example 1 and the corresponding average TU game $\tilde{v}$, we have

$$\mathrm{PGV}_1(\tilde{v}) = \frac{51}{18}, \quad \mathrm{PGV}_2(\tilde{v}) = \frac{44}{18}, \quad \text{and } \mathrm{PGV}_3(\tilde{v}) = \frac{42}{18}.$$

To sum up, we have seen that different generalizations of the Public Good value for TU games or the normalized Public Good index for simple games to the class of (non-trivial) (j, k) simple games, including axiomatizations, are possible. As anticipated, e.g., in Freixas (2012), a power index for simple games can admit more than one reasonable extension for (j, k) simple games. From our personal point of view, Theorem 1 provides the most convincing variant. But this may be just a matter of taste or might depend on the application. The question of the public good properties of the proposed values is not touched at all. As done in Courtin and Tchantcho (2020) for $(j, 2)$ simple games, other power indices based on *Riker's Size Principle* (Riker, 1962, p. 32) may be treated similarly.

References

Allingham, M. G. (1975). Economic power and values of games. *Zeitschrift für Nationalökonomie, 35*(3–4), 293–299.

Courtin, S., & Tchantcho, B. (2020). Public Good indices for games with several levels of approval. *The BE Journal of Theoretical Economics*, *20*(1).

Dubey, P. (1975). On the uniqueness of the Shapley value. *International Journal of Game Theory, 4*(3), 131–139.

Freixas, J. (2005). The Shapley-Shubik power index for games with several levels of approval in the input and output. *Decision Support Systems, 39*(2), 185–195.

Freixas, J. (2012). Probabilistic power indices for voting rules with abstention. *Mathematical Social Sciences, 64*(1), 89–99.

Freixas, J., & Zwicker, W. S. (2003). Weighted voting, abstention, and multiple levels of approval. *Social Choice and Welfare, 21*(3), 399–431.

Freixas, J., & Zwicker, W. S. (2009). Anonymous yes-no voting with abstention and multiple levels of approval. *Games and Economic Behavior, 67*(2), 428–444.

Haradau, R., & Napel, S. (2007). The potential of the Holler-Packel value and index: A new characterization. *Homo Oeconomicus, 24*(2), 255–268.

Holler, M. J. (1982). Forming coalitions and measuring voting power. *Political Studies, 30*(2), 262–271.

Holler, M. J., & Li, X. (1995). From public good index to public value: An axiomatic approach and generalization. *Control and Cybernetics, 24*, 257–270.

Holler, M. J., & Packel, E. W. (1983). Power, luck and the right index. *Zeitschrift für Nationalökonomie, 43*(1), 21–29.

Kurz, S. (2020). Which criteria qualify power indices for applications? : A comment on "the story of the poor Public Good index." *Munich Social Science Review, 3*, 185–198.

Kurz, S., Moyouwou, I., & Touyem, H. (2021). Axiomatizations for the Shapley-Shubik power index for games with several levels of approval in the input and output. *Social Choice and Welfare, 56*(3), 569–594.

Riker, W. H. (1962). *The theory of political coalitions* (Vol. 173). Yale University Press.

Public Economics, Politics and Responsibility

Why Did Japan Attack Pearl Harbor?

Ronald Wintrobe

1 Introduction

Why did Japan attack pearl harbor? It was obvious they could not win a war with the US. Did they not anticipate that the US would react? Did they think they had the capacity to win the war? The most common answer is that the attack was simply not rational. In this paper I first briefly [in the next section] consider the most obvious common explanations for the attack and suggest that they are all unsatisfactory. I then [Sect. 3] advance an explanation based on the idea that Japan in the first half of the twentieth century was what I call a "quasi theocracy"–a regime where rule is divided between a religious and a civilian authority. Because decision making was divided between the religious and secular authorities, and because there was no formal separation of individual values from the state, an inversion of power occurred whereby decisions from the top were in effect led by the military, and decisions by senior military were influenced by the actions of their subordinates, in a process which was dysfunctional for the whole. The third section outlines this explanation. Some other examples of quasi theocracy may be modern Iran, or Turkey under the Ottoman Empire. Section 4 discusses these other potential applications, especially Iran, very briefly. Section 5 then describes the decision making processes leading up to the Pearl Harbor attack. Section 6 concludes the paper.

R. Wintrobe (✉)
Western University, London, Canada
e-mail: rwintrob@uwo.ca

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_15

2 Some Common Explanations for the Attack

We can begin with Roberta Wohlsetter's statement of her viewpoint, which is also the common one in the literature, in her magisterial book *Pearl Harbor: Warning and Decision:*

> The relation between this material [the facts about the relative military strengths of Japan and the US at that time] and the decision to take on the US as an opponent is simply not explicable in rational terms (p. 352)…..

Or, in the words of Iguchi Takeo, in a recent book outlining the decisions leading up to the attack from a Japanese point of view "Japan's decision to go to war lacked rationality" (Takeo, (2010) p. 93).

Let's be more specific, since this question and the issues surrounding it are so important to understanding what happened then, and what may be happening again in the contemporary politics of other countries like Iran, Israel and the US. The economic theory of conflict identifies several circumstances that can lead to war (Blattman, (2022); Fearon, (1995) Jackson and Morelli (2007)). The most common of these is that each of the parties miscalculates the likelihood that it can win, either because it miscalculates its opponent's strength and mistakenly thinks its forces are superior, or underestimates it's opponent's resolve (Fearon, (1995).

Did the Japanese government miscalculate the strength of US forces? Here is Wohlsetter again:

> The Japanese did have the material for making long range predictions….their assessment of our war potential in aircraft manufacturing, shipbuilding, and rate of training of the necessary crews was much more accurate than our own for 1941, 1942 and 1943….
>
> ….Japanese and American estimates of the risks to the Japanese were identical for the large-scale war they had planned, as well as for the individual operations. What we miscalculated was the ability and willingness of the Japanese to accept such risks. As Ambassador Grew had said, "National sanity would dictate against such an event, but Japanese sanity cannot be measured by our standards of logic" (Wohlsetter, p.354)

Another rationality-based explanation for the attack is that there were irreconcilable conflict of interests between the US and Japan. Now, of course, there was a deep conflict, over China and particularly Manchuria. But if Japan and the US had competing interests there, why could they not just divide up the territory (the colonies), as the Western governments did in the Middle East and elsewhere after World War 1?

Another possibility, less rationalistic in tone, but still possibly satisfying in a way, is simply that the Japanese are particularly warlike. However, it is worth remembering that Japan was totally peaceful for the 250 years of the Tokugawa shogunate, which lasted until US Admiral Perry sailed to Japan and demanded they "open up".

A fourth possible explanation is simply that Japan was a dictatorship, and it is well known that dictatorships make war more frequently than democracies do. In particular, attacking the US could simply have been part of a xenophobic "greater Asia" strategy. But if Japan was a dictatorship, who was the dictator?

The Japanese political philosopher Masao Maruyama describes the wartime Japanese government as revealed by the evidence presented at the Tokyo "war crimes" trials after the war:

> …during the period covered by this indictment fifteen separate cabinets rose and fell in Japan .. in the rise and fall of these many cabinets composing the government of Japan there were twenty -one prime ministers, thirty foreign ministers, twenty--eight home ministers, nineteen war ministers ... Rather than to establish an agreement or common plan or conspiracy, the purpose of which was to dominate the world or any other objective, the evidence definitely reveals the absence of leadership or of a centralized group committed to a common design or purpose of any kind (Masao Maruyama p. 86)

What about the emperor? Wasn't he a dictator? Could Japan have gone to war be a case of "leadership bias", an argument elaborated formally in Jackson and Morelli, (2007), in which a country's leaders expect positive net gains, while this may not be so for the members of their side?

The Emperor of Japan, the core repository of Japanese spiritual values, was not merely God's representative, as in other theocracies like the Papacy or Calvin's Geneva, he himself *was* divine, in theory the direct descendant of Jimmu, the first emperor and the son of the Goddess Ameratsu (in fact he could only trace his lineage back to the fourteenth century).

Bix (2000) describes Hirohito's upbringing. He observes that Hirohito's Ethics teacher Sugiura was an ultranationalist. Sugiura taught Hirohito that in foreign countries the relationship was determined by power but in Japan "the emperor rules the people without power. Benevolence has been planted so deeply in the minds of the people that ….the people joyfully submit themselves to the emperor "…. for the emperor to lay burdens on his subjects was entirely natural because they existed to sacrifice themselves for him, not the other way round" (Bix, (2000), p. 65).

But the emperor did not *rule*; a civilian administration ruled in his name. So Japan over this period was not a true theocracy, like the Papacy during the era of the Papal States where the Pope was God's representative and ruled directly. There were two areas where the Japanese emperor had sole authority: the military, and education. But this was not written in stone, and sometimes the military during this period thought that they should be completely autonomous, with no authority over them. In fact, it seems that the powers of the emperor waxed and waned, depending on individual force of personality and on circumstances. Hirohito's father Taisho was very weak, for example, and his powers accordingly became circumscribed. As Bix (2000) notes: "(As a child) Hirohito witnessed the practise of reducing the sphere of imperial court assent to the smallest possible extent, so that neither his father the emperor nor he himself would need to express the "imperial will (Bix (2000) p.130)". Even Hirohito's grandfather, the illustrious Meiji, never made a military decision. But later, sometimes Hirohito became powerful behind the scene, as discussed later. His precise role in the Pearl Harbor attack is complex and is discussed in detailed in Sect. 5.

3 The Argument Here: Quasi Theocracy

(i) The concept of quasi theocracy

A quasi theocracy is a regime where rule is divided between a civilian administration and a religious authority. Japan after the Meiji restoration is the most obvious case, but there are others: modern Iran, and possibly Ottoman Turkey.

What might be called *the quasi - theocrat's quandary* is that, on the one hand, the quasi theocrat may grow bored with a limited ceremonial role and want more power. But If he does that, he risks undermining other political authorities and emboldening bureaucrats to act in his name. Most important, his position is not inviolable and if he tries to augment his power he could be removed. To try to rule openly and directly may result in not ruling at all.

To put it differently, the emperor is divine, but only if he is the right emperor! So, for example, Hirohito had a brother, Prince Chichibu, who appeared much more charismatic than he did. Chichibu was next in line to the throne, and Hirohito received secret reports about his activities from his steward (Bix p.179).

So if the emperor does try to augment his power it has to be in secret so that.

(a) if things go wrong, he can't be blamed, and
(b) he is not accused of usurping his role

In Hirohito's case, on the one hand while his official role was a that of a kind of shrine, as Buruma (2004) puts it, [1] a robot, or a mouse, on the other hand sometimes he appeared to be powerful, and made and unmade cabinets and prime ministers behind the scenes. This dual role led to a psychological dilemma. Of course, from the economic point of view one could suggest that the quasi theocrat can solve his quandary by simply accumulating power to the point where the marginal benefit of more power is equal to its marginal cost, but what a formulation like that would miss is that the choice question involved implies a dual identity: ruler, or shrine? As a ruler he may maximize power, as a shrine he merely presides over ceremonial occasions and the like and should not be, or even appear to be interested in power.

The next step in understanding the working of a regime where the leadership is divided, and the two leaders compete for power, one religious and one secular, is to see the implications of this divided leadership for the position of the bureaucracy.

The basic implication is simple: Because the leadership is divided, the bureaucracy has more power. Agencies can use this power to advance their interests. As a consequence our basic and central prediction about a quasi theocracy is that it tends to result in an inversion of power, with the bottom (the bureaucracy) controlling the top (the civilian and religious leadership). The most obvious beneficiary of the division of authority is the military.

[1] The emperor could not be held responsible for the war because he was a mere shrine.

(ii) The military

In Japan in particular the military had more power than is typical of other types of regimes because after Meiji it was formally subordinate only to the emperor, not to the civilian authorities. So the most obvious beneficiary of this division was the military. What did they want? More power, and a bigger budget. What is the best way to get that? Prepare for war.

Now, I have argued elsewhere (Breton and Wintrobe, (1975), Wintrobe (1997)) that, in general, the Niskanen (1968) model is not a good theory of bureaucracy because:

(1) Bureaucrats generally are mobile, and therefore need not budget maximize to get more income, prestige, or power
(2) The information distortions that they use to accumulate power can be controlled in various ways. For example, governing politicians can utilize redundant bureaus, external data checks, "spies" within the bureau,[2] and so on to do this.

But the military are a special case: they are neither mobile nor easily controlle. So budget maximization may indeed be a good theory in this particular case. Indeed, Niskanen himself worked for an organization that tried to control the defense department and it was from that experience that he formulated his theories (see Niskanen (1975)).

The problem of controlling bureaucratic maneuvering in the defense department continues in the United States today: in particular Bob Woodward's book on the Obama administration *Obama's War*, (2010) provides fascinating material on the difficulties Obama had. For example, the US military insisted they had to have 40,000 more troops in Afghanistan, that a "war game" showed that 20,000 would not do. But no war game was ever conducted, as the current President, then Vice- President Biden found out and reported to the president. Nevertheless the military continued to insist that there were only three options: 40,000 troops, 40,000 troops and 40,000 troops…..(Woodward, 2010).

It is interesting that in Obama's own account of his presidency (Obama, 2020), Biden emerges as the strongest opponent of the military's demands for more resources. That opposition seems to have continued under his own presidency, and led to his abrupt withdrawal from Afghanistan. One could speculate that this well known opposition on Biden's part may have led to him being deceived into misunderstanding the consequences of the disastrously precipitous nature of the withdrawal.

Returning to Japan, it is important to emphasize that the military there was not a monopoly. The Army and the Navy competed for funds and their leaders did not necessarily share the same views. Lower levels also tried to make policy, especially by taking action in their own hands (as in the Manchurian incident, described shortly).

But they all schemed and clamored for bigger budgets (Bix, (2000), Buruma (2004), Prange and Goldstein and Goldstein (1981), Finer, (2002).

[2] See Breton and Wintrobe (1975) or my survey of bureaucracy (Wintrobe (1997) for details of these criticisms of the Niskanen model.

This thesis–-that control passed from the top to the bottom in Japan in the1930's[3]–is common in the historical literature (e.g., Finer, (2002), Maruyama, (1969), Bix, (2000), Buruma (2004), Takeo, (2010), Wohlsetter (1962). But those same sources often suggest that what drove Japan to war was the "concern for the preservation of Japan as a whole –" the *kokutai* (the Japanese body politic–see, for example Buruma (2004) or Wohlsetter (1962)).

For example Wohlsetter (1962) summarizes her argument on this point as follows:…"the decision for war with the US was not chosen. The decision for war was rather forced by the desire to avoid the more terrible alternatives of losing status or abandoning the national objectives (Wohlsetter (1962), pp. 352–353)."

Now, the argument here is exactly the reverse of this proposition: it is because no one had the responsibility and the power to make decisions on behalf of the whole that each party made decisions which served its own interests and the result was disastrous for the whole.

(iii) Japanese spiritual values

There were no wars of religion in Japan to result in the separation of church and state as in Europe. As a result morality was defined by the *kokutai,* i.e., whatever pleases the Emperor, or whatever helps the state. It was not separate from the state.

In *any* theocracy, there is the idea that individual values and morality are not separate from the state, but embodied in it. This gives individuals a basis for thinking that their actions should embody the values of the state, that is, spiritual values which embody the "mission" of the nation.

In a *quasi* theocracy, the fact that spirituality is part of the state aids the lower level bureaucrats in using their own "morality" to act on their own—in effect it gives them more power. The military, whose duties are always especially in need of a moral justification, is particularly vulnerable to this incentive. In turn the judges who judge their actions are also free to disregard the law if they feel it conflicts with morality. So the propaganda of *hakku ichio*, (holy war) in spiritualizing the state, led further to the inversion of power. Consider, for example, the remarks by General Tojo. He stated that *if Japan were to compromise on withdrawal from China, he would never again be able to sleep facing the Yasakuni shrine* (quoted in Maruyama, (1969) pp.134–5).

Another example is General Matsui, who argued against making concessions to the Americans and British along the following lines:

> If we were now to settle the [China] Incident by compromising with England and America and co-operating with the Americans, how would we be able to face the myriad spirits of the war dead? Ultimately, it is for the sake of the myriad spirits of the war dead that we are so adamantly opposed to any compromise with America. (quoted in Maruyama, (1969) p.134)

Lest the reader think this reflects a particularly Japanese point of view, it is worth pointing out that Bob Woodward has made the same argument with respect to American involvement in Afghanistan (Woodward, (2010). He argues that since many

[3] In the 1920's, military budgets fell, democracy gained ground, policy towards China was conciliatory, and in 1930, the Japanese Prime Minister signed the Treaty of London without the consent of the Army. It was only after this that the senior military officers in the army began their drive for power (Finer, The Man on Horseback, p. 45).

American soldiers have fought bravely and lost their lives in that war, the Americans have to stay in to win, otherwise they dishonor that memory.

(iv) Agenda transformation

The concept of agenda control is familiar in Public Choice theory, especially since the work of Mckelvey, (1976). Here we advance a different version, based on the Japanese experience, of "agenda transformation", in which the bottom can exert control over the decisions of the top. The theory is best illustrated with a simple example. A military official takes an aggressive action which promotes war where no war has been declared. This action changes the government's choice space from.

X_0, x_1, and x_2 to $x_0 - L$, x_1 and x_2.

Where L is the loss of retreating to the original status quo position, which has been changed from x_0 to $x_0 - L$.

So the decision space for the government is changed from

Make War if

(1) $U(SQ) < p\,U\,(SQ + G) + (1 - p)\,U\,(SQ - LW) - C$

To:

(2) $U(SQ - L) < p\,U\,(SQ + G) + (1 - p)\,U\,(SQ - LW) - C$

Where

SQ = status quo

G = the gains from war if victorious, neglecting the costs of conflict

LW = the losses from war if defeated, not including the costs of conflict

C = the costs of conflict

L = the losses if the government backs down

As is apparent from inspecting the equations, the left hand side of (2) is less than that in (1), so with the new choice space available to the government in (2), the government is more likely to go to war. In this way, lower level military personnel can take control of the government's decision making and possibly get the government todo what they want it to do. An illustration from the Japanese experience follows.

(v) Illustration: the Manchurian Incident

On September 18, 1931, near Mukden (now Shenyang) in southern Manchuria, a section of railroad owned by Japan's South Manchuria Railway was dynamited. The Imperial Japanese Army, accusing Chinese dissidents of the act, responded with the invasion of Manchuria, leading to the establishment of Manchukuo the following year. While the responsibility for this act of sabotage remains a subject of controversy, the prevailing view is that Japanese militarists staged the explosion in order to provide a pretext for war.

It is widely felt (see Takeo, (2010) in particular) that the establishment of Manchuko was crucial in leading the Japanese to first make war against China, and later the United States. In Sect. 5 we will see how the occupation of Manchuria played into the decision to attack Pearl Harbor.

(vi) Summary

1. In a quasi theocracy like Japan the bureaucracy has more power.
2. The military are the most important beneficiaries of this inversion of power, because their peculiar structure allows and incentivizes them to use it to increase their budgets through information distortion, agenda control and agenda transformation.
3. "Spiritual values" enhance the autonomy and power of the military.
4. But senior levels within the military or other bureaucracy are themselves vulnerable to pressure from below.

4 Other Examples? Ottoman Empire, Iran

Two other possible examples of quasi theocracy are the Ottoman Empire and contemporary Iran. We discuss them here, beginning with a very brief discussion of the Ottoman case.

The early sultans of the Empire were powerful but the later ones, according to Samuel Finer's *History of Government*…were as "incompetent as the first 10 were magnificent". What kept the Empire going? "The bureaucracy" he says. He also points out that the empire was at war 7 years out of 10, indeed, he asserts that "the empire lived by war and off war" (Finer (1997), p. 1165). But he does not advance the argument made here that the regime was so warlike *because* the bureaucracy was so competent and powerful.

In Turkey, as in Japan and Iran, the Wars are "holy wars" -*ghazi* (Turkish). In Japanese, the most common concept was *hakku ichiu* or "eight corners of the world under one roof"- usually interpreted as the benign rule of the Emperor. The structure of decision making at the upper levels appeared similar in all of the cases of the emperor, the sultan and the ayatollah. Thus:

The emperor attended meetings of the civilian cabinet but normally he did not speak.

The sultan hid behind the curtain while his cabinet deliberated. He observed the proceedings but he did not speak.

The ayatollah is thought to have more power than this -for example, it was Khomeini himself who made the decision to use the Guardian of the Jurist as a constitutional model. But the ayatollah's position is not inviolable:

> the overlapping and complicated power structure of the Islamic Republic was intentionally designed in 1979 to obfuscate lines of authority and ensure no single entity became powerful enough to bend to foreign pressure, resulting in a myriad of power centres, *none with supreme authority*. The Supreme Leader Ayatollah Khamenei has the most power….but..he may not act in isolation like Iran's last monarch, Mohammad Reza Shah Pahlavi.. (Tarzai (2009) p. 2) italics added).

Indeed, perhaps the most well known of American scholars of the Iran revolution, Nikki Keddie, in her book *Roots and Results of Revolution: An Interpretive History of Modern Iran,* (1981) insists that "Iran is "not a dictatorship". It has never been

clear to me what she meant by this, but I think we can understand a bit more if we recognize the competition for power between the religious and secular branches of government.

Perhaps enough has been said on the subject of Iran to provide some support for looking at that regime as a quasi theocracy, i.e., a regime where there is competition between the religious and secular arms of the government, and that the main beneficiary of this competition is the military. Let us now return to the case of Japan and see how this competition played out in that case.

5 Pearl Harbor

(i) The approach to pearl harbor: after the Manchurian incident

After the Manchurian Incident, Prime Minister Inukai Tsuyosh attempted to revive amicable relations with the Nationalist government of Chiang Kai Shek and keep a tight rein on the insubordinate Japanese military. However, it proved to be his undoing, and on 15 May 1932 he was assassinated by fanatic military officers. (Takeo, (2010), p. 44).

The sentences passed to those responsible for these and other acts were often absurdly light, their acts of terrorism justified by "patriotic motives" (Morton and Olenik 181, Takeo, pp. 44–5). On February 26, 1936, junior Imperial Way officers tried to take over central Tokyo, assassinating the finance minister and attempting to assassinate the Prime Minister (but killing his brother in law by mistake). The Navy was called in to restore order, and they did, but the price was that the military demanded that new cabinet officers be approved by the army and navy ministers, themselves serving officers.

After the assassination, the new Prime Minister was Hirota, who raised the military budget as the army demanded. But neither he nor his successors could control the troops in China. No war with China was ever declared but war began anyway with the China Incident of July 7, 1937, a small and accidental battle which seems to have served as the precursor for the Japanese invasion of Manchuria. War continued but it was inconclusive, and the United States demanded that the Japanese cease the aggression against China. A crucial resolution was agreed to at the imperial conference of 6 sept 1941, which stated:

> If by the early part of October there is no reasonable hope of having our demands agreed to…we will immediately make up our minds to get ready for war against America and England and Holland

In October 1941 a meeting was held to discuss this resolution between Prince Konoe, PM General Tojo, the ministers of War and the Navy, the Foreign Minister, and Lieut Gen Suzuki, Director of the Planning Board. General.

Suzuki recalled the meeting at the Tokyo trials (from Maruyama p. 88, see also Bix, Buruma):

> It became quite clear as the result of this conference where the thorny question lay. The Navy really thought that the war with America was impossible but did not desire openly to say so. The Army did not necessarily desire war, but vigorously objected to the withdrawal of troops from China. The Foreign Minister was firmly of the opinion that without consenting to the withdrawal of the Armed Forces from China the negotiations with America offered no prospect of success. The only way for the Prime Minister to avoid war was, therefore, either to make the Navy declare its real intentions, or to make the Army understand the unexpressed intentions of the Navy and agree to the withdrawal of the armed forces. I saw that the Prime Minister was in a predicament because personally he felt himself unequal to the task of persuading the Navy or the Army (IMTFE, no. 333, p. 35206, 12 Dec 1947, quoted in Maruyama, (1969), p. 88)

Did the (senior) military want war with the west? Admiral Yamamoto (the planner of The Pearl Harbor attack) told PM Konoe: "Japan can successfully challenge the US for a year at most" (Buruma, (2004) p.115). Americans stopped selling aviation fuel and scrap metal to Japan in January 1940, oil in 1941. Then there was the Hull note: in November 1941 US foreign minister Cordell Hull wrote to General Tojo (the new Prime Minister) demanding withdrawal from Indochina and China but he made no mention of Manchuria. General Tojo presented this to the government as an ultimatum, which it was not (Bix, (2000), Buruma, (2004)), and insisted that the Japanese could not yield.

And if the Americans and the British refused to back down? General Tojo:

> sometimes people have to shut their eyes and take the plunge". (quoted in Buruma, (2004), p. 119)).

But there is still the question of Hirohito, the Emperor. If he didn't support the attack, it is not obvious that the military could have decided it without him, or at least, over his formal opposition, since he was, after all, their commander.

(ii) the decision to attack pearl harbor

On the actual decision to attack, and Hirohito's role in it, there is a famous and startling account which is told both in Bix and in Gordon Prange and Goldstein's At Dawn We Slept: The Untold Story of Pearl Harbor. I quote from the latter:

Hirohito was not bent on war and informed Premier Konoye on Sept 5, that "he wished to question the chiefs of the Army and Navy General Staffs at the forthcoming imperial conference—an unprecedented step" (Prange and Goldstein (1981), 208). Konoye, who "may have been genuinely horrified at the idea of the Emperor's taking an active role at an imperial conference" (Prange and Goldstein (1981) 209) suggested that he summon the chiefs for a private audience instead. At that meeting, the night before the imperial conference, Hirohito told the two officers that he wished diplomacy to be emphasized and was unconvinced by their assurances that it was……Hirohito then asked Sugiyama "how long he thought hostilities would last in case Japan and the United States went to war. The general estimated that operations in the South Pacific could be disposed of in about three months. Thereupon the blood rose in Hirohito's face and he answered Sugiyama in an unusually loud tone: "As War Minister at the outbreak of the China Incident, you asked me to approve sending Army troops there, saying that the Incident would be settled in a short time. But has it yet ended

after more than four years? Are you trying to sell me the same thing again?" (Prange and Goldstein and Goldstein (1981), p. 209, italics added; see also Bix, (2000)

At the imperial conference the next day, "the real question was not, Shall we fight? but When shall we fight?" (Prange and Goldstein and Goldstein (1981), p.210) and Hirohito decided to speak up, reading a short poem by his grandfather Meiji, "The Four Sides of the Sea".

Methinks all the people of the world are brethren, then
"Why are waves and winds so unsettled nowadays?"

explaining that he was "striving to introduce into the present the emperor Meiji's ideal of international peace" (Prange and Goldstein (1981) p. 211). The Prime Minister Konoye recorded that "Everyone present was struck with awe" (Prange and Goldstein (1981), p. 211). But after this incident Hirohito appeared to drop his opposition, and at the next meeting he didn't raise any further objections (Prange and Goldstein, Bix).

Was Hirohito's behavior irrational? It is true that after the war Hirohito was forced to renounce his divinity. "I am not divine" he said in a radio address which was the first time many people heard him speak. He was not indicted as a war criminal (at MacArthur's insistence) and reigned until 1992.

(iii) Why did the U.S not anticipate the attack on Pearl Harbor?

This is the subject that more than anything else has engaged American scholars and produced bi, important books such as those by Prange and Goldstein and Goldstein (1981) and Wohlsetter (1962). Again, it may be useful to quote Roberta Wohlsetter:

> The fact of surprise at Pearl Harbor has never been persuasively explained by accusing the participants, individually or in groups, of conspiracy or negligence or stupidity. What these examples illustrate is rather the very human tendency to pay attention to the signals that support current expectations about enemy behavior (Wohlsetter (1962), p.392)

Some interesting examples of these are worth mentioning:

1. Even after the war, the judges at the Tokyo trials could not grasp the nature of the Japanese regime, they looked for a single dictator and chain of command, as in Germany (see Maruyama, (1969)).
2. The Japanese conducted a war game to see what would happen after Pearl Harbor. America was modelled as a single entity, Germany and Italy as one, but they modelled the Japanse as several entities, the Army, Navy, etc. and looked to see what would happen depending on who got the upper hand. (Wohlsetter)
3. There were errors. For example, the message sent on the morning of Dec 7 by the Japanese, and intercepted by the US Navy, as many were by the US' codebreaking system known as "MAGIC", and which referred to the final rupture of negotiations with the Americans, was translated as

> Relations between Japan and England are not in accordance with expectations

However the message actually said:

> Relations between Japan and England and the United States are on the brink of catastrophe

Since the Japanese are normally inclined to understatement, it might have been useful to have had the correct translation.

6 Conclusion

In this paper I have focused on the decision by the Japanese government to attack Pearl Harbor in 1941. But the argument made here is more general than that, as I have also emphasized, and in this conclusion let me stress three more general points:

1. Just because it's irrational for a country as a whole to make war doesn't mean it won't happen. War could occur because although it is irrational from the point of view of the country as a whole, it may be rational from the point of view of individual decision makers within that country whose interests are more narrow and who can force a decision to go to war through agenda transformation or through other means. In particular this may be true of quasi theocracy, and this leads me to my second conclusion.
2. The type of regime that I have labeled quasi theocracy in this paper may be more dangerous to international stability than democracy or dictatorship
3. Finally, the paper points to a flaw in models of bureaucracy which assume monopoly bureaus. The various military bureaus in Japan competed fiercely leading up to the war, and their behavior was extremely destructive for the polity. So one implication of the Japanese experience in World War II is that competing bureaus cannot be counted on to check each other and expose the information distortions of their competitors. The opposite might occur, i.e., competition might produce more dysfunctional behavior rather than reduce or eliminate it. So competition among bureaus may be worse than monopoly from the point of view of the welfare of the polity. Clearly, more research needs to be done on the theory of competitive budget maximization.

References

Arendt, H. (1951). *The Origins of Totalitarianism*. Harcourt Brace.
Arendt, H. (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. Viking Press.
Bix, H. (2000). *Hirohito and the Making of Modern Japan*. Harper Collins.
Blattman, C. (2022). *Why We Fight*. Viking Press.
Breton, A., & Ronald, W. (1975). The equilibrium size of a budget maximizing bureau: A note on Niskanen's theory of bureaucracy. *Journal of Political Economy, 83.*
Breton, A., & Wintrobe, R. (1986). The bureaucracy of murder revisited. *Journal of Political Economy*
Buruma, I. (2004). *Inventing Japan 1853-1964*. New York: Modern Library
Fearon, J. D. (1995). Rationalist explanations for war. *International Organization*
Ferrero, Mario, & Wintrobe, Ronald (Eds.). (2009). *The Political Economy of Theocracy*. Palgrave.

Finer, S. (2002). *The Man On Horseback*. Transaction Publishers
Finer, Samuel. (1997). *The History of Government from the Earliest Times.* Oxford University Press.
Jackson, M. O., & Morelli, M. (2007). Political bias and war. *American Economic Review. 97*, 1353–1373
Keddie, N. (1981). *Roots and Results of Revolution: An Interpretive History of Modern Iran.* Yale University Press
Mackay, Robert, & Weaver, Carolyn. (1981). Agenda control by bureaucrats in a multi-bureau setting. *Public Choice, 37*, 447–472.
Mckelvey, Richard. (1976). Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory, 12*, 472–482.
Maruyama, Masao. (1969). *Thought and Behavior in Modern Japanese Politics.* Oxford University Press.
Morton and Olenik, 1970 Morton & Olenik. (1970). Japan: Its History and Culture. New York: McGraw Hill
Niskanen, W. (1975). Bureaucrats and Politicians. *Journal of Law and Economics, 18*(3).
Niskanen, W. (1968). The peculiar economics of bureaucracy. *American Economic Review*
Niskanen, W. (2007). *Bureaucracy and Representative Government.* Aldine Transaction
Obama, Barack. (2020). *A Promised Land*. Crown.
Prange and Goldstein, Gordon W., & Goldstein, D. M. (1981). *At Dawn We Slept: the Untold Story of Pearl Harbor.* New York: Penguin
Romer, T., & Rosenthal, H. (1979). Bureaucrats versus voters: On the political economy of resource allocation by direct democracy**.** *Quarterly Journal of Economics, 93*(4), 563-587
Takeo, I. (2010). *Demystifying Pearl Harbor: A New Perspective from Japan.* Translated by David Noble. Tokyo: International House of Japan
Tarzai and Piece, 2009Tarzai, A. (2009). The Iranian Puzzle Piece: Understanding Iran in the Global Context. Quantico, Virginia: Marine Corps University Press
Wintrobe, R. (1997). Modern Bureaucratic Theory. In D. Mueller (Ed), *Public Choice: A Handbook.* New York: Cambridge University Press
Wintrobe, R. (1998). *The Political Economy of Dictatorship*. Cambridge University Press.
Wohlsetter, R., & Harbor, P. (1962). *Warning and Decision.* Stanford University Press.
Woodward, B. (2010). *Obama's Wars.* Simon and Schuster.

Inflated Expectations of Democracy: Towards a Systematic Explanation

George Tridimas

This essay is based on my 8 March 2021 talk to the Munich Adam Smith Seminar convened and organized by Manfred J. Holler, in cooperation with Gregor Berz, Florian Rupp, and Eva Wolf. I thank the participants for a stimulating discussion during the presentation. I also wish to thank an anonymous Referee for comments and suggestions offered on a previous draft. The usual disclaimer applies.

It is true that we are called a democracy, for the administration is in the hands of the many and not of the few.

Pericles, 431 BCE, in Thucydides' Funerial Oration

Democracy is government of the people, by the people, for the people.

Abraham Lincoln, 1863

There never was an election without a party.

W. Bagehot, 1873, The English Constitution

Democracy is necessarily and inevitably party government.

H. Kelsen, 1929, Vom Wesen und Wert der Democratie

Democracy is the worst form of Government except for all those other forms that have been tried from time to time.

Winston Churchill, 1947

1 Introduction

The opening quotations aim to capture that the ideal of democracy has been the same in ancient and modern times (Pericles and Lincoln separated by 2,300 years), that in modern times democracy has morphed into party competition and party government (Bagehot and Kelsen) and that democracy does not offer panaceas (Churchill).

Present Address:
G. Tridimas (✉)
Department of Accounting Finance and Economics, Ulster University Business School, York Street, Belfast BT15 1ED, UK
e-mail: G.Tridimas@ulster.ac.uk

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_16

According to Finer (1999:1568) "a democracy can be defined as a state where political decisions are taken by and with the consent, or the active participation even, of the majority of the People." In modern democracies, elections to choose candidates for government office using some form of majority voting provide a peaceful method of dealing with conflicts about who has the power to take political decisions.[1] In addition, and rather informally, democracy itself is seen as both an ideal, and a mechanism expected to produce ideal outcomes. Yet, such popularizations confuse the issue of how to select a government based on individual preferences and whether the outcome of government policies is good or moral on some (hopefully) unanimously held ethical principle.[2] These confusions are at the root of misunderstandings and inflated expectations about democracy ultimately leading to disappointments which feed anti-democratic tendencies.

During the long line of history, democratic government has been the exception rather the rule. It is only over the last two hundred years or so that democracy in the form of representative government slowly emerged and flourished first in the traditional "West" and gradually in several other countries around the world. In the process, expectations about the performance and outcomes of democratically elected governments have been inflated. "What often many people seem to equate with democracy is popular participation in the political process, individual freedom, equality, honest government, and economic prosperity" (Munger, 2005: 125). Whether these ideals are compatible with each other, force inevitable trade-offs, or, more importantly, are obtainable through practical political arrangements is overlooked. As a result, dissatisfactions with democratic outcomes abound and solemn warnings that democracy is in crisis are heard with increasing frequency.[3] Indeed, democratic "backsliding" in for example, contemporary Hungary, Poland, Turkey and Venezuela, or the 6th of January 2021 invasion of America's Capitol by mobs of supporters of Donald Trump after losing the presidential election are some of the latest, most obvious signs of the fragility of democracy.

Democrats worry about many threats including attacks by foreign autocrats, the risk that democracy degenerates to the rule of the mob leading to anarchy that offers authoritarians the opportunity to seize control, and about incumbents eroding the democracy from within. More systematically, the existential threats to democracy can be distinguished between extrinsic or environmental and intrinsic or foundational. Extrinsic threats originate from attacks by foreign non-democratic powers or terrorists, economic failures propagated by domestic and international factors, technological changes which strengthen centralization and control over individuals, and domestic occupiers of office undermining liberties and democratic institutions. Intrinsic causes originate from possible design faults in the institutions of democratic

[1] "In all democratic regimes the principal officers of government are chosen through competitive elections in which the bulk of the population can participate" (Huntington 1991:580).

[2] It bears noting that even though the issue of deciding the mechanism for selecting a government is a positive question, a modicum of normative principles still applies to it insofar we would like the mechanism to take account of criteria like inclusiveness.

[3] See among others Berman (2019); Levitsky & Zibatt (2018); Przeworski (2019); Runciman (2019).

governance. Obviously, in practice the two types of causes may work in tandem to imperil democracy. A large and growing volume of research has investigated the role of the extrinsic factors; in particular, how economic, social, and ideological determinants and their evolution may precipitate crises in democracy. However, the focus of the present inquiry is the set of intrinsic causes, the innate shortcomings of the institutions of democracy that make it vulnerable.

The paper is structured as follows. Section 2 summarizes the advantages of democracy. Section 3 describes the grievances against modern democracy. Section 4 uses the perspectives of institutional economics and social choice theory to explain the analytical basis of dissatisfaction with democracy. It investigates how the following factors may affect attitudes towards democracy: voting as a mechanism of aggregating individual preferences; the challenge of undefined voter preferences and imperfect information of voters; the impossibility of a collective choice equilibrium as argued by social choice theory; the agency nature of political representation where voters elect politicians who are organized in parties; the political inequality arising from differences between voters and differences between voters and candidates for political office; and finally, the ethical neutrality of majority voting outcomes and simultaneously the need to uphold democratic norms. Section 5 concludes.

2 The Enduring Appeal of Democracy

The word democracy is derived from the Greek term *democratia*, which means people have power, a form of governance invented and first practiced by the Ancient Greeks.[4] This form of participatory democracy of fully enfranchised male-only citizens flourished for almost two hundred years in the period 510–322 BCE but came to an end with the rise of centralized Hellenistic kingdoms and then the Roman conquest of Greece finalized in 146 BCE. Modern representative government is today's democracy. The idea and practice of some form of democratic governance emerged after a long interval, in the seventeenth century, following momentous events in England. However, representative government as first established in late eighteenth century USA was conceived in opposition to the Classical Greek democracy. Landemore (2020:3) captures this:

> The representative systems we now call 'democracies' were indeed initially intended as anything but democratic. Instead, they were initially the product of a liberal-republican, rather than strictly democratic, value system. By 'liberal,' I mean here deriving from an ideology primarily concerned with protecting the inalienable rights of individuals against

[4] See Cartledge (2016) for a recent account of the influence of ancient Greek ideas of democracy on eighteenth century French Enlightenment thinkers (especially Montesquieu, Rousseau and Voltaire) and nineteenth century British thinkers (Burke, Paine, Grote and John Stuart Mill). For a dissenting view that before its rise in Classical Athens, an early form of democracy was practiced in the Americas, precolonial Central Africa, Mesopotamia, and India see Stasavage (2020); however, he uses the less stringent condition of seeking consent as the defining condition for democracy (for an extended discussion of Stasavage see Tridimas, 2022).

> the encroachment of governments, including popular governments. By 'republican,' I mean an ideology where the ideal of non-domination of the individual trumps the ideal of popular rule (though those two ideals can perhaps be reconciled). This priority of liberal-republican commitments and goals over purely democratic ones was compatible with giving the people some say over the choice of rulers, but not as clearly compatible with the ideal of popular rule per se.

Elections then became a key element of representative government. Although at first only a small, enfranchised elite had access to government, the extension of suffrage over the nineteenth century produced the institutions and practices of the modern democracy.[5] Nowadays, a regime is democratic when it has regular, competitive elections, without fraud or manipulation, and with universal adult suffrage.[6]

Democracy has both intrinsic and instrumental value. The intrinsic value is that democracy serves the appealing principles of equality, individual freedom, and citizen sovereignty. That is, under democratic government citizens respectively have equal say manifested through one-person one-vote, which in turn leads to some form of majority rule[7]; citizens are free to choose on private and public matters; and they are the final arbiters over every matter. Further, as Runciman (2019) put it, by respecting individual preferences democracy puts the dignity of the individual at center-stage. The instrumental value of democracy is that Western-type democracies have experienced long-term benefits in the form of material prosperity, social stability, and peace (although formally establishing the causal link between democracy and economic prosperity has proved problematic).

More specifically, the appeal of modern representative democracy is that it legitimates the government (Manin, 1997). Government is, by definition, coercive since the ruled are obliged to do as they are told. If one accepts that individuals are obliged only by what they have consented to, the ruler is legitimate when the ruled have

---

[5] The economics literature on the emergence of modern democracy and specifically the extension of the franchise is immense. Formal models highlighting the significance of redistribution are found in among other the work of Acemoglu & Robinson (2006); Boix (2003); Ansell & Samuels (2014). Congleton (2011) emphasises the gradual as opposed to revolutionary rise of representative government and the supporting intellectual currents. Berman (2019) focuses on conditions conducive to the establishment of liberal democracy in Europe and the obstacles it has confronted. Stasavage (2020) analyses the development of democratic and autocratic modes of state governance from pre-historic times to today.

[6] This is the definition used by Freedom House. Freedom House's term "electoral democracy" differs from "liberal democracy" in that the latter also implies the presence of a substantial array of civil liberties. See Freedom House www.freedomhouse.org Freedom in the World (2013:29). Empirically, there is a strong positive correlation between democracy and both economic and political freedoms, but the correlation does not prove causation. The inference here is that "a non-democratic regime could fully realize liberal freedoms. Similarly, a democracy could run roughshod over its citizens' civil and economic liberties" (Brennan, 2018:336).

[7] See Buchanan & Tullock (1971) for the optimal majority rule using economic calculus. Space considerations prevent a detailed discussion of the type of majority rule applied, simple or qualified, and related questions of the voting rule, first-past-the post, or proportional, and the formulas used when adopting the latter.

consented.[8] In the modern electoral democracy, the consent of the ruled is obtained by electing rulers in a fair election. In competitive, fair, and recurring elections the people judge and select those who are the best to govern. Equality of citizens is achieved by offering everyone an equal opportunity to consent through the rule of one-person one-vote.

Achen & Bartels (2016) distinguish between two approaches regarding the attractiveness of modern democracy. The first, termed the "folk theory" or populist ideal of democracy, maintains that ordinary citizens determine the policies of democratic communities. According to this view people as voters are wise, informed and engage in politics. As a result, what the majority wants becomes government policy after voting for representatives (or in referendums). Rather optimistically, the folk theorem assumes a high level of voter information and sophistication. The second approach, termed the leadership selection model, praises democracy for the opportunity it offers voters to punish politicians who fail to deliver good policies. This view is based on a retrospective theory of voting, where election outcomes depend on voter approval of actual performance of incumbent politicians. Worryingly, although perhaps informationally less demanding than the folk theorem, the retrospective view models citizens as driving "the automobile of state simply by looking in the rear-view mirror". However, Achen & Bartels present evidence which rejects both these views of democracy. They find that the great majority of citizens pay little attention to politics, denying therefore the validity of the folk theorem. Similarly, they offer evidence that voters are myopic with limited ability to judge politicians correctly; indeed, they examine cases where the latter inability was so large that voters punished incumbents for things outside their control, like acts of nature.

3 A Litany of Complaints Against Democracy

According to Przeworski (2019) a crisis in democracy occurs when it does not generate an outcome required for its survival. The list of grievances against modern democracy leading to a crisis makes for long and distressing reading. The complaints relate to failures to satisfy both the intrinsic and instrumental values. A summary of those goes as follows.

In the first instance, modern representative democracy has become inaccessible and unresponsive to citizens. "Many people feel neglected. Their views seem to count for little and their representatives often appear uninterested in hearing them out." (Runciman, 2019:141). Although with universal suffrage citizen participation is broad, it is thin and discontinuous. Voters have virtually no input on policy making and their preferences are recorded only in periodic elections which relate to choosing political representatives rather than policies. Citizens then are more like passive consumers rather than active participants in shaping the policies which affect their

[8] This reflects the principle "*quod omnes tangit, ab omnibus tractari, et approbari debet*" ("QOT"): that which touches all shall be dealt with and approved by all (Manin 1997; Stasavage 2020).

interests. What is more, twenty-first century representative democracy has empowered an elite of politicians, their associates, and domestic and international business interests opening an ever-growing gap between political leaders and ordinary citizens as governments are losing the will and ability to curb corporate power and its excesses. This is further exacerbated as the balance of power has shifted in favour of the executive resulting in weaker parliaments, while national governments have lost out to remote and less accountable supra-national governance organizations.

Second, disappointing economic developments have shaken the faith to democracy's ability to deliver continuing economic prosperity. After the so-called "golden age of capitalism" of the period 1948–73, in the more recent past, mature western democracies have experienced lower than expected growth, market volatility, loss of manufacturing jobs, and increasing income inequality. Even more alarming are survey findings reporting that respondents think that their children will be worse off than themselves.[9]

Dissatisfaction with representative democracy is manifested in the declining voter turnouts in election after election, the falling membership rates of political parties (although other socio-economic factors have contributed to this outcome too), falling trust in the institutions of democracy (especially the ability of elections to resolve problems), declining confidence in elected politicians (of both government and opposition parties), as well as increasing incidences of public disorder, and the rise of populism of the left or right hue characterized by an authoritarian streak. Populism is driven by the charge that democracy has been stolen from the people by the elites. Populism tends to deny the legitimacy of established parties and attacks them as undemocratic and unpatriotic threatening the very survival of democracy.

4 A Choice-Theoretic Explanation of the Disappointments from Democracy

Democracy is an institution, that is, a rule constraining individual actions, for arriving at collective choices. Social choice theory seeks to identify how a group of individuals may reach a collective decision which binds them all. It has shown that collective decision mechanisms, methods of aggregating individual preferences on issues of public interests to reach an equilibrium outcome, are subject to severe problems.[10] This section borrows from the economics of institutions and social choice theory to develop a systematic understanding of the disappointments with democracy and its hallmark, majority voting.

[9] See Przeworski (2019) for a review of surveys showing a decline of support for democracy.

[10] See Mueller (2003) for a book-length analysis of the relevant topics.

4.1 Aggregation of Voter Preferences

The exercise of power in public policy making binds all individuals and is thus coercive. For example, the decision to go war applies and constrains those who are pro-war as well as those who are against it. Similarly, and less dramatically, traffic regulations apply to all, those who wish to speed and those who prefer to go slowly. However, different individuals have different preferences about policy outcomes, and a clash of interests inevitably arises, resulting in winners and losers. Democracy is one of several different types of institutions which can be used to arrive at a decision obliging all participants. It does so by avoiding violent conflict. If unanimity is not possible or too expensive to operate, a type of majority rule is used to aggregate individual preferences.[11] Voting then becomes the key element of the minimalist concept of democracy. The one-person one-vote condition equalizes the input from each voter into the collective choice process. But then majority voting gives rise to the following conundrum. With aggregation over a large number, each voter realizes that her vote has little, if any, influence over the outcome.[12] This may well be a cause of frustration with democracy: having been told that in a democracy she is in control, the voter realises that her input to the decision making is of no consequence. The result is what Przeworski (2019:202–3) calls "the inescapable fact that each one of us must be ruled by someone else and being ruled must entail policies and laws we do not like."

4.2 Unarticulated Voter Preferences and Incomplete Voter Information

A prerequisite for making a collective decision on public policy issues is that voters have well-defined preferences, so that they make meaningful choices which are then aggregated through the voting mechanism to derive a social choice outcome. But do voters have well-defined preferences for public policy issues? There is a distinct and very real possibility that voters are preoccupied with their daily struggles in their own lives leaving little time and other resources to get informed and form preferences about political outcomes.

For Caplan (2007:14) voters are irrational in the sense that in addition to an interest in wealth maximization, they hold "cherished views, valued for their own sake; acting on these beliefs is costly but confers psychological rather than material benefits". He examines four types of such beliefs that people may hold, anti-market, anti-foreign, make-work (prosperity comes from employment rather than production), and pessimistic (the economy is doing worse than it actually is). Achen & Bartels (2016:32) also claim that many citizens "do not have meaningful beliefs…

[11] Ibid.

[12] See Brennan (2018).

on issues". Empirical evidence on the presence of framing effects, where different but equivalent wording of a question on the same issue leads respondents to give different answers, question the empirical validity of the assumption that citizens have well-defined preferences. If this is true, neither does it make sense to expect a well-defined democratic outcome, nor should we be surprised that voters are disappointed by whatever outcome prevails.

Economists who model individual preferences as pre-existing (rather than asking how they are formed) are more worried about the incomplete information used by voters to vote and, more generally, make public policy decisions. The problem here is that political issues may be too complex for most voters to comprehend and thus engage in politics in a meaningful way. Downs (1957) talked about the rational ignorance of voters. For an individual voter, the costs of obtaining information on issues of public interest and actively participating in the political process are high and certain. On the contrary, reflecting the issue of aggregation mentioned before, the probability that her vote is decisive to swing the voting outcome to her preferred policy is infinitesimally small so that an individual vote makes no difference to the outcome. Hence, for a voter facing a high and certain cost but a small expected benefit from voting, it is rational to remain ignorant of what is involved in the political process, unless she expects a large personal benefit. If this is true, only those with direct stakes will participate in politics and political outcomes, and government policies will reflect the balance of power between special interest groups with possible detrimental effects for the welfare of the mass of voters.

Lacking the relevant information voters also lack the ability to judge issues and politicians advocating different policies, which implies that they may be attracted by simplistic solutions. Voters can neither evaluate whether "times are good or bad" nor can they estimate whether incumbent politicians or external circumstances are responsible for the good or bad times. If this is the case both the folk theorem and the retrospective theory of voting lose their salience as ways to explain and evaluate democracy.[13]

However, this may not be the end for democracy. Political entrepreneurs wishing to win election have strong incentives to provide information to voters about the benefits of their preferred policies and against those of their rivals. So, voters are certain to get a modicum of information since they choose politicians typically organized in political parties. This type of delegation from voters to politicians raises a new host of problems for representative democracy which are examined in Sect. 4.4. The next sub-section looks at the question of whether a social choice outcome exists upon assuming that preferences are defined, and information is complete.

[13] Note the seeming paradox: In the age of internet and social media, a proliferation of news sources makes it harder for voters to know what is really going on. Invoking limited-information and bounded-rationality, the paradox is explained by the argument that people gravitate towards what they want to hear, so no one is any the wiser.

4.3 Impossibility of a Social Choice Equilibrium

In his pioneering work Arrow (1951) demonstrated that no process for aggregating individual preferences can produce a social ordering that satisfies five, seemingly mild conditions, and an endless cycle arises where no policy proposal commands a majority against all other proposals. The problem of cycling (or failure to establish a social choice equilibrium) has motivated a voluminous literature examining both the normative question of how democratic processes ought to work and the positive question of how they work in practice.

If voting is on a single dimension issue of public policy (for example, the size of the tax rate) and individual preferences are single-peaked (graphically, utility has an inverse-U shape), simple majority voting leads to a stable collective choice equilibrium, the median voter outcome. That is, the median voter's ideal point wins against all other policy points in pairwise voting (Downs, 1957). However, when voting is on a multi-dimensional decision space (for example, the size of tax rates and the allocation of the revenue between, say, civilian and military expenditure), or when preferences are not single-peaked, or voter choices are influenced by factors unrelated to the issue voted upon (that is, "irrelevant alternatives"), a social choice equilibrium may not exist.

The cycle is avoided when some individual preferences are excluded, or the choice is left to a single individual (a dictator), or a particular arbitrary structure is imposed on the content, or order, of options offered. Neither exclusion, nor dictatorship, nor arbitrary agenda may be appealing ways to secure a collective choice outcome. For those who want to defend democracy as a system that leads to good government based on the will of the people a sad realization dawns: The general will of the voters may not be discovered by a procedure which counts individual preferences, such as voting.[14] On this account, an idealized view of democracy gives rise to inflated expectations from the mechanisms of democratic choice.

Whether the theoretical prediction of cycling is significant for the stability of the political system is challenged by Miller (1983). Invoking pluralist theory, which considers the society as containing clusters of individuals with different preferences about public policies, he challenges the significance of cycling for the stability of the democratic system. "Pluralist political theory identifies certain patterns of political preferences (reflecting certain social and economic structures) as promoting the "stability" of democratic political systems… the preference patterns identified by social choice theory as leading to stable choice are essentially those identified by pluralist theory as destabilizing for the system (p:734). Further, "Commonly in pluralist democracies, there is indeed fairly regular alternation of winners and losers in successive elections. It is very *important to try to understand what brings about this alternation.* The most obvious answer is that there are substantial shifts in the distribution of political preferences over time" (p.744, emphasis in the original).

[14] This insight may also, at least, partly explain the disappointment with the performance of newly democratized countries after the collapse of communism in Europe and in Asia and the Arab Spring.

In practice, policy choices are made within institutions of decision-making which among other things define the size of the franchise, that is, who has the right to vote; the electoral rule, that is, simple or qualified majority, first-past-the-post, proportional representation, division of the electorate into geographical constituencies size; agenda-setting powers, namely, what issues are to be decided through voting, who has the right to propose legislation, the order of taking votes, and veto power. Such institutions, or "rules of the game" lead to what is known as structure induced equilibriums, that is voting outcomes which depend on the type of institution adopted (see Shepsle & Weingast, 2012 for a recent review). But institutions affect incentives which affect efficiency, and determine rights which affect distribution of income and wealth, so the induced collective choice outcome may be thought as arbitrary or unfair by some individuals, who may then voice loud concern about the operation of democracy.

4.4 The Agency Nature of Political Representation

4.4.1 Politicians as Representatives of Voters

Designing and implementing policy by politicians, that is, political entrepreneurs who are specialists in issues of public interests and are selected by the citizens for their proposed policies, brings benefits from division of labour and specialization. However, these benefits come at a cost. Voting for political representatives who are better informed than voters and are granted discretionary powers, creates an agency relationship where the principal-voter delegates to the agent-politician to act on the former's behalf to achieve desirable effects.

There are two notions of representation (Manin, 1997; Stasavage, 2020). The first is the descriptive (or statistical, or mirror) representation, where the representative and the represented share the same identity; this form dominated pre-modern types of representation. In this model, representatives operate under mandate; they reflect the interests of the represented and decide in ways the represented would themselves have decided had they taken the decision themselves. Second, trustee representation, where the representative is a trustee with full powers ("*plenam potestatem agendi*") independent of the represented and makes her own judgements concerning the interests of the whole nation.

The delegation relationship from the citizens to political leaders raises a new host of problems for representative democracy. Agency generates the well-known problems of adverse selection, where the principal cannot observe important exogenous characteristics or contingencies affecting its welfare, and moral hazard, where the principal is unable to observe and monitor the actions of the agent to make sure that they are consistent with and conducive to fulfilling the terms of the appointment. In the context of political representation, adverse selection takes the form of doubts voters may have about the kind of persons who are seeking political office, and raises questions relating to the motivation, honesty, and competence of politicians. Moral

hazard problems relate to the risk that political office holders may exploit information advantages and discretionary powers so that they pursue their own interests instead of the interests of the represented.

Such problems may be detrimental for democracy. First, they may cause short-termism. That is, politicians interested in winning electoral contests, pander to ephemeral pressures and ignore adverse long-run implications of quick-fix policies. Further, if controls on the executive power are weak, elected office holders may behave autocratically so that democracy is at risk. In addition, if, as popular sentiment often has it, politicians are lying (for example, by making false charges of election fraud), they undermine public confidence in elections. But when citizens do not trust the electoral process, they often lose faith in democracy itself. In the light of this discussion, we may conclude that representative democracy is subject to the conundrum that representative government requires that the government be accountable to the representatives of the citizens and simultaneously the representatives must be held into account (Lord & Pollak, 2010: 975).

4.4.2 Voting for Political Parties

In representative democracy, the agency relationship between voters and politicians is not so much a personal relationship but one mediated through political parties. Political economy research explains that there are several, in truth complementary, reasons for the emergence of political parties.[15] Politicians form political parties to pass legislation, as no individual legislator can pass a bill, a problem of collective action (Aldrich, 1995), and to agree on a set of policies to avoid cycles, a problem of collective choice impossibility (Aldrich ibid, Jackson & Moselle, 2002). At the same time, imperfectly informed voters vote for political parties, since parties are organizations which aggregate ideologically similar candidates and resolve various information and commitment problems. Specifically, political parties as "brands" that outlast politicians provide low-cost information signals about the preferences of candidates (Levy, 2004), discipline politicians to credibly commit to policies in the long-run interests of the electorate (Alesina & Spear, 1988), and exploit economies of scale in collecting and disseminating information on policy issues (Aldrich, 1995; Osborne & Tourky, 2008). Further, parties are formed, financed, and otherwise backed by interest groups, like industrialists, labour, farmers or environmentalists, to pursue the policy demands of their constituents (Bawn et al., 2012).

When democratic politics is conducted by party competition a new range of issues arise that may lead to disappointment with democracy. In the first instance, and pursuing the intuition gained from acknowledging the principal-agent relationship of representative democracy, in party elections voters do not decide policy. Rather, they vote for politicians who then decide policy. Second, given the central role of parties in modern representative democracy, politicians are a hybrid of the representation

[15] See Tridimas (2019) for a review and a contrast with the absence of political parties from the Classical Greek direct democracy.

types described above, neither do they have full mandate from the electors nor full independent powers, but party discipline forces them to be faithful to their party. Representative democracy takes the form of competition between parties rather than democracy within parties (like market competition found between firms rather than within firms). Hence, the relation of voters to policy making is at best tenuous.[16]

The third most alarming issue is that voters vote for vague party platforms which bundle policies for different issues. Inevitably, casting a single vote to decide policies on a multitude of issues restricts voter choices and the voting outcome may no longer register the breadth and intensity of preferences on each issue decided. Calling referendums for specific policy issues offers the opportunity to unbundle issues and allow voters a direct input into decision making (Mastusaka, 2004). Nevertheless, a referendum may neither reveal popular preferences nor bring the electoral outcome to the median voter equilibrium when agenda setting, in the sense of the right to ask the referendum question, remains uncontested in the hands of the government (see Tsebelis, 2002 and Altman, 2019 for relevant analytical perspectives), as arguably shown by the UK 2016 Brexit referendum.

With the great majority of citizens paying little attention to politics, political outcomes and government policies will reflect the balance of power between special interest groups, an outcome which repudiates both the folk theorem and the retrospective view of democracy. Achen & Bartels (2016) argue that in this circumstance, election outcomes are mostly just erratic reflections of the current balance of partisan loyalties in a given political system. For them political outcomes are explained by what they call the "group theory of democracy." According to this view citizens are members of social groups. The social identities and group attachments determine their political loyalties. The upshot is that group and partisan loyalties, not policy preferences or ideologies, are fundamental determinants of election outcomes. The inference here is that voting is expressive rather than instrumental. That is, voters vote for a particular party not as an instrument to pursue policy gains, but because of the psychological benefits from expressing their identity and class, which in turn are reflected by political parties. Hence, regardless of its instrumental usefulness, voting expresses some aspect of the voter's beliefs, values, ideology, identity, or personality (see Brennan & Hamlin, 1999; Hamlin & Jennings, 2011; Hillman, 2010; Aachen & Bartels, 2016). But when political parties represent not just policy differences but ways of life, identity politics arrive. Identity politics fuels populist frustrations.[17] Runciman (2018) interprets the rise of identity politics as a signal of voter frustration. This frustration is born from the fact that voters are only asked to consent to power, while they wish to control power. Having achieved the right to be listened to (through universal suffrage), twenty-first century voters seek to be heard too.

[16] Holding party primaries for selecting party leaders is a step towards party democracy and is supported by the folk theory of democracy, but the problem of imperfect voter information is not resolved: Committed party members may select candidates with extreme views.

[17] Bukodi & Goldthorpe (2021) draw attention to the existence of evidence consistent with the view that low social status individuals support populist parties with their anti-elitism slant, and they are more authoritarian.

4.5 Political Inequality

Regarding equality, the fundamental creed of representative democracy is "one-person one-vote." Democracy and voting are supposed to equalize power: the rich and the poor have one vote apiece. However, in practice we come across grave discrepancies from this principle, as at least two types of inequalities loom large. First, there is economic inequality, where unlike ordinary citizens, rich individuals, large corporations, and business associations fund the campaigns of politicians and lobby them between elections in exchange for privileged treatment and economic rents. Hence, organized policy demanders get what they want at the expense of the unorganized groups. Second, there is voting inequality, which arises when constituency size and electoral laws end up (intentionally or not) assigning different weights to different voters, so that some votes count more than others, as for example in marginal seats where voters swing between competing parties and safe seats where a party seems to obtain repeatedly unassailable majorities. As a result, a party (or a presidential candidate) may end up winning a plurality of the electoral vote but not enough seats for a majority in parliament (or electoral college) which prevents it from assuming office. In such cases equality of voters is more elusive than real.

Representative democracy confronts a second inequality, that of candidates for office. Obviously, candidates have unequal chances of being elected to office because of their different policy platforms, but the concern here is that, in addition, the probability of electoral victory depends on non-policy attributes, what is termed as valence. Specifically, voters may treat different candidates differently on any criterion they like (Manin, 1997). In the hustings, each politician emphasizes what makes her superior to other competitors and more suitable for public office. These include factors such as character, looks, achievements in sports or the battlefield, that is, issues which may animate voters one way or another, but are outside politics.

Yet another cause of inequality between candidates is differences in the possession of financial resources. Informing voters about policies and candidates is costly. Such costs are barriers to entry to the political marketplace and distort the level-playing field of the political contest. Richly endowed candidates have higher probabilities of electoral victory. Since it is easier to collect funds from a small number of big contributions, politicians tend to rely on 'big' donors (corporations, rich persons, or groups of unionized workers) for funding their campaigns and reward them accordingly after winning the election. Again, a gap opens between voters and the governing elite.

It then follows that elections produce office holders with characteristics 'superior' to those of voters, that is, persons who on balance are wealthier, better networked, and more educated. The reliance on the party machine for funding and promotion tends to favour career or professional politicians. This difference is further compounded by an additional trend, where professional politicians rising through the party machine without experience of work at anything else but party organizations, political think-tanks, and public relations. Although this type of labour division and specialization increases efficiency, it also increases the gap between the rulers and the ruled. That the elected office holders comprise an elite in comparison to ordinary electors-voters

has long been recognized by none other than Aristotle in his Politics (book 6), who considered elections as "aristocratic" meaning that they give an advantage to the well-born who have the means to become known and train in public speaking to win the support of the citizens. In this vein, Manin (1997) points out that, etymologically, the words election and elite have the same root.

4.6 Democracy and Ethical Behaviour

4.6.1 Ethical Neutrality of Aggregation of Preferences

To attribute intrinsic and instrumental values to democracy does not automatically translate to democracy securing ethical outcomes. As already said, majority voting is a mechanism of aggregating individual preferences when politics is a competition for hearts and minds, but it is fallacious to demand that the mechanism produces an ethical outcome. To put it another way, democratic procedures may be a fair mechanism of aggregation of preferences, and majority voting may legitimate the outcome, but the outcome may not necessarily be morally legitimate, that is, good from a normative perspective. Democracy is a framework, not an outcome; it is not about discovering what is "good" for the society. As Munger (2005: 120–21) put it "the majority preference is simply what most people happen to think. It has no moral force, other than as a means of resolving disputes."

In truth, and depending on what is considered as moral, the majority voting outcome may be sinister. For example, the majority may vote to oppress a minority of citizens, or abuse human rights, or elect politicians who pursue opportunistic foreign policies and even fight disastrous wars. Examples from the ancient and the modern democracies abound.[18]

Protection of individuals from the abuses of the majority calls for restricting the freedom of the majority to do as it likes. To prevent abuses, constitutions grant individuals rights and empower courts to annul policy measures which violate such rights.[19] It is for this reason that the modern democracy is known as "liberal democracy".[20] However, endowing courts with the power to annul measures decided by elected governments raises different and sometimes acute problems. Politicians frustrated that the judiciary blocks them may accuse the judges of political interference and partisan bias, which then stirs anti-elitist sentiments.[21]

---

[18] It also implies that majoritarian institutions may fail to secure peaceful politics in societies deeply divided along racial, religious, or cultural lines.

[19] See Mueller (2001) for the rationale of conferring rights in an uncertain environment. See Broman & Vanberg (2021) and the literature therein for a choice-theoretic analysis of the creation of independent courts.

[20] Fawcett (2018) offers a recent account of the ideology and practice of liberalism.

[21] The same applies to other non-majoritarian institutions, like politically independent central banks entrusted with price stability. More generally, Przeworski (2019:4) warns that "The problem of adding adjectives to 'democracy' is that not all good things must go together".

4.6.2 The Significance of Democratic Norms

Even though democratic procedures and majority voting are best seen as ethically neutral mechanisms, the operation and performance of such mechanisms is not independent of the ethics of those who operate the mechanisms. As already emphasised, democracy is a set of rules of the game about the power to decide policy. As the economics of institutions alerts us, imperfect information about citizen preferences, unanticipated technological change and external shocks, render formal constitutional rules about how to decide policy incomplete contracts which are subject to competing interpretations. Hence, the success of a democracy depends crucially on all competing players holding and accepting shared democratic norms of behaviour, what may be called a democratic culture.

Probably, the first necessary (but not sufficient) condition for the latter is a common sense of belonging to a cohesive society. It is under this circumstance that citizens are convinced that arriving peacefully at collective choices is worth the effort and accept the coercion implied by being ruled. But more than that is warranted.

Examining the importance of a democratic culture Levitsky & Ziblatt (2018:101) write.

> These rules or norms serve as the soft guardrails of democracy, preventing day-to-day political competition from devolving into a no-holds-barred conflict. Norms are… shared codes of conduct that become common knowledge within a particular community or society—accepted, respected, and enforced by its members. Because they are unwritten, they are often hard to see, especially when they're functioning well.

They stress two (closely related) fundamental norms of conduct. First, mutual toleration, the principle that if all players play by the same rules, they all accept that each one has an equal right to exist, compete for power, and govern. Second, forbearance, the exercise of patient self-control and restraint, which may even go as far as refraining from using the full extent of permissible institutional prerogatives.[22] For Levitsky & Ziblatt, intense partisan polarisation breaks the basic norms of mutual tolerance and forbearance. Writing about the contemporary US, they castigate both the Republicans and the Democrats party (but the former to a greater degree) for corrosive behaviour. Well before the storming of Capitol Hill, they had charged President Donald Trump for breaking the norms of American democracy by amongst other things derailing independent investigations, not separating private and public affairs, and lying to the electorate. To these, false accusations of stolen elections were then added.

In the UK which lacks a codified constitution, the unwritten norms are referred to as the "good chaps" model of government, a term coined in 1985, where "a good chap knows what a good chap has to do and doesn't need to be told" (Hennessy, 1994:205). In what is probably the pinnacle of political principal-agency and deference to their political superiors, under the good-chaps-government citizens "have

[22] Levitsky & Ziblatt (ibid: 107) "Think of democracy as a game that we want to keep playing indefinitely. To ensure future rounds of the game, players must refrain from either incapacitating the other team or antagonizing them to such a degree, that they refuse to play again tomorrow".

trusted politicians to behave themselves. [Citizens] have long assumed that those who rise to high office will be 'good chaps', knowing what the unwritten rules are and wanting to adhere to them, even if doing so might frustrate the attainment of their policy objectives, party political goals, or personal ambitions—the argument being that 'good chaps' (of different sexes) know where the undrawn lines lie and come nowhere near to crossing them" (Blick & Hennessy, 2019:4–5). However, the ructions following the 2016 Brexit Referendum have cast serious doubt whether everyone involved in politics is a "good chap", bound by codes of honour and reliable to do the right thing when running the country. In this connection, one of the most telling events was the 2019 attempt by the Prime Minister Boris Johnson to prorogue (suspend) Parliament amid Brexit negotiations, which was struck down by the UK Supreme Court. "This episode was regrettable both because it represented an improper attempt to circumvent Parliament; and because it drew the monarchy, its powers and functions into party political controversy" (ibid:10).[23]

All in all, the success of an ethically neutral mechanism relies on the actors being ethical.

5 Conclusions

The present essay inquired what deficiencies in the institutional framework of democracy may lead to disillusionment with democracy that in turn may make citizens receptive to non-democratic alternatives. Democracy is a blessing because it allows peaceful change in government through elections. But democracy is also "a curse… because there seems to be an irresistible urge to attach moral force to the will of the majority, when in fact it is nothing more than what most people happen to think." (Munger, 2012:65). If democracy, in the sense of people's power, is seen as an ideal leading to discover the general will, and if it is acknowledged that different people have different and contradictory preferences and values, then democracy generates inflated expectations and democratic outcomes are bound to disappoint, perhaps deeply, many people some of the times. If, less ambitiously, democracy is seen as a peaceful mechanism of aggregating preferences by using some form of majority rule, then expectations are more restrained, innate shortcomings of democracy are acknowledged, and become the subject of positive investigation. The present paper was written in the latter spirit.

The foundational defects of majority voting identified here were divided into the following, partly overlapping, categories: (1) problems with aggregation of individual preferences; (2) incomplete voter preferences and imperfect information; (3) the impossibility of a social choice equilibrium; (4) the principal—agent nature of voting for political representatives; (5) inequalities across voters and across politicians; and (6) the lack of ethical dimension of voting as a method of aggregation of preferences,

[23] See Tridimas & Tridimas (2020) for a public choice analysis of the ruling of the Court.

and simultaneously the need to adhere to moral norms of behaviour for the survival of democracy.

This acknowledgment injects a dose of realism to what we should expect of democracy. Democracy possesses enviable advantages, but it is not a flawless governance system. Having walked the tortuous path to achieve it, its survival cannot be taken for granted. To use a famous line from Demosthenes, the great fourth century Athenian orator and committed democrat, "it often proves harder to keep than to win".[24] Despite all this, we can also end by striking note of optimism from Runciman's (2019:186). "Even if democracy is often bad at coming up with the right answer, it is good at unpicking the wrong ones."

References

Achen, C. H., & Bartels, L. M. (2016). *Democracy for realists*. Princeton University Press.
Acemoglu, D., & Robinson, J. A. (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.
Aldrich, J. (1995). *Why parties?* University of Chicago.
Alesina, A., & Spear, S. E. (1988). An overlapping generations model of electoral competition. *Journal of Public Economics, 37*, 359–379.
Altman, D. (2019). *Citizenship and contemporary direct democracy*. Cambridge University Press.
Ansell, B. W., & Samuels, D. J. (2014). *Inequality and democratization. An elite-competition approach*. Cambridge University Press.
Arrow, K. (1951). *Social choice and individual values*. New Haven: Yale University Press (2nd ed., 1963).
Bawn, K., Cohen, M., Karol, D., Masket, S., Noel, H., & Zaller, J. (2012). A theory of political parties: Groups, policy demands and nominations in American politics. *Perspectives on Politics, 10*, 571–597.
Berman, S. (2019). *Democracy and dictatorship in Europe: From the Ancien Régime to the present day*. Oxford University Press.
Boix, C. (2003). *Democracy and Redistribution*. Cambridge: Cambridge University Press.
Blick, A., & Hennessy, P. (2019). *Good chaps no more? Safeguarding the constitution in stressful times*. The Constitution Society.
Brennan, G., & Hamlin, A. (1999). On political representation. *British Journal of Political Science., 29*, 109–127.
Brennan, J. (2018). Democracy and freedom. In D. Schmidtz & C. E. Pavel (Eds.), *The Oxford handbook of freedom* (pp. 336–348). Oxford University Press.
Broman, B., & Vanberg, G. (2021). Feuding, arbitration and the emergence of an independent judiciary. *Constitutional Political Economy* (forthcoming).
Bukodi, E., & Goldthorpe, J. H. (2021). Meritocracy and populism. Is there a connection. *United Kingdom in a changing Europe*. Retrieved from 12–07–2021 https://osf.io/preprints/socarxiv/qgrkf/.
Buchanan, J. M., & Tullock, G. (1971). *The calculus of consent*. University of Michigan Press.
Caplan, B. (2007). *The myth of rational voter. Why democracies choose bad policies*. Princeton University Press.
Cartledge, P. (2016). *Democracy. A life*. Oxford University Press.

[24] Demosthenes, First Olynthiac, 1.23, available at https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0070%3Aspeech%3D1%3Asection%3D23.

Congleton, R. D. (2011). *Perfecting parliament: Constitutional reform and the origins of western democracy*. Cambridge University Press.
Downs, A. (1957). *An economic theory of democracy*. Harper & Brothers.
Finer, S.E. (1999). *The history of government. Empires, monarchies and the modern state* (Vol. III). Oxford: Oxford University Press
Fawcett, E. (2018). *Liberalism. The life of an idea* (2nd ed.). Oxford University Press.
Hamlin, A., & Jennings, C. (2011). Expressive political behaviour: Foundations, scope and implications. *British Journal of Political Science, 41*, 645–670.
Hennessy, P. (1994). Harvesting the cupboards: Why Britain has produced no administrative theory or ideology in the twentieth century. *Transactions of the Royal Historical Society, 4*, 203–219.
Hillman, A. L. (2010). Expressive behavior in economics and politics. *European Journal of Political Economy, 26*, 403–418.
Huntington, S. P. (1991). How countries democratize. *Political Science Quarterly, 106*, 579–616.
Jackson, M. O., & Moselle, B. (2002). Coalition and party formation in a legislative voting game. *Journal of Economic Theory, 103*, 49–87.
Landemore, H. (2021) *Open Democracy. Reinventing Popular Rule for the 21st Century*. Princeton: Princeton University Press.
Levitsky, S., & Zibatt, D. (2018). *How democracies die.* Penguin Random House LLC.
Levy, G. (2004). A model of political parties. *Journal of Economic Theory, 115*(2), 250–277. S0022053103002540, https://doi.org/10.1016/S0022-0531(03)00254-0
Lord, C., & Pollak, J. (2010). Representation and accountability: Communicating tubes? *West European Politics, 33*, 968–988.
Manin, B. (1997). *The principles of representative government*. Cambridge University Press.
Matsusaka, J. G. (2004). *For the many or the few: The initiative, public policy, and American democracy*. University of Chicago Press.
Miller, N. R. (1983). Pluralism and social choice. *The American Political Science Review, 77*, 734–747.
Mueller, D. C. (2003). *Public choice III*. Cambridge University Press.
Mueller, D. C. (2001). The importance of uncertainty in a two-stage theory of constitutions. *Public Choice, 108*, 223–258.
Munger, M. (2005). Nineteenth-century voting procedures in a twenty-first century world. *Public Choice, 124*, 115–133.
Munger, M. (2012). Voting methods, problems of majority rule, and demand-revealing procedures. *Public Choice, 152*, 61–72.
Osborne, M., & Tourky, R. (2008). Party formation in single–issue politics. *Journal of the European Economic Association, 6*, 974–1005.
Przeworski, A. (2019). *Crises of democracy*. Cambridge University Press.
Runciman, D. (2019). *How democracy ends.* Profile Books.
Shepsle, K. A., & Weingast, B. R. (2012). Why so much stability? Majority voting, legislative institutions, and Gordon Tullock. *Public Choice, 152*, 83–95.
Stasavage, D. (2020). *The decline and rise of democracy*. Princeton University Press.
Tridimas, C. P., & Tridimas, G. (2020). Is the UK Supreme Court rogue to un-prorogue Parliament? *European Journal of Law and Economics, 49*, 205–225.
Tridimas, G. (2019). Democracy without political parties: The case of ancient Athens. *Journal of Institutional Economics, 15*, 983–998.
Tridimas, G. (2022). Book Review: The decline and rise of democracy, by David Stasavage. *Constitutional Political Economy, in Press*. https://doi.org/10.1007/s10602-022-09367-9
Tsebelis, G. (2002). *Veto players: How political institutions work*. Princeton University Press.

Electoral Competition, Political Promises and the "Responsible Party Government" Hypothesis

Benoît Le Maux

> To be responsible is to be accountable for one's actions, to be trustworthy; politically, to be answerable. Responsibility also implies an element of predictability. In the context of political parties, responsibility has always been associated with issues. One may conclude that party responsibility implies two things. The first is that parties take clear positions on a series of issues that concern voters. Second, parties, to the extent that they are returned to office, seek to implement their positions in the policy-making process. (Wray, 1981)

1 Introduction

The belief that political markets are not necessarily efficient is shared by many political scientists and political economists. In 1950, a supplement of the American Political Science Review, "Toward a more responsible two-party system", was published to bring insight into the weaknesses of the U.S. two-party system (Committee on Political Parties, 1950). The reason for the report lied in the argument that American parties were "*loose associations that lacked coherent philosophies or ideologies or programmes and that were unable to act in the disciplined fashion necessary for effective governance*" (Cronin, 2019). Meanwhile, Black (1948) and Downs (1957) were deriving a set of formal results suggesting that political parties have little interest in holding contrasted opinions; they will instead strategically choose to be moderate in order to get elected, as initially emphasized in Hotelling (1929). A few years later, Buchanan and Tullock (1962) were addressing a set of governmental inefficiencies, assuming that politicians are self-interested, as any agent of the private sector. Since then, the literature has flourished in the field, addressing the question of whether parties were "responsible".

The purpose of this essay is to review the literature on the two main issues that characterize the *responsible party government hypothesis*: (1) whether political parties

B. Le Maux (✉)
University of Rennes 1, CREM-CNRS, Condorcet Center, Rennes, France
e-mail: benoit.le-maux@univ-rennes1.fr

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_17

have clear and distinct positions and (2) whether they hold to their pledges once elected (Wray, 1981). It aims to bring together and summarize the knowledge that has been gained in political economy and related disciplines since the first developments. This piece of work can also be useful to new scholars in the field, who might feel lost in the myriad of studies that have been published in the past 70 years. Obviously, given the limited amount of pages, this review cannot be exhaustive and will, more simply, focus on the main challenges of the literature. We invite the reader to read Persson and Tabellini (2000), Mueller (2003), Padovano (2013), Jan (2015), Duggan and Martinelli (2017), De Donder and Gallego (2017) and Winer and Ferris (2022) for extended views, among others.

We will see that many models of electoral competition posit that political candidates implement their announced platform once elected or, at least, are incentivized to do so. Candidates who wish to have a career in politics may resist opportunities to make false promises in order to maintain a profitable relationship with the electorate. Voters can thus rationally vote for their favorite candidate knowing that promises will be fulfilled—that is to say, democracy would be a powerful force inducing desirable political outcomes. In this view, parties would be "*responsible*" in the sense that they are "*ideologically polarized*" (Bernhardt et al., 2009), thus providing the electorate with enough choice, and their behavior could be easily predicted by *"consistently projecting what they have done previously"* (Downs, 1957).

This optimistic view that parties are responsible, however, holds under highly restrictive assumptions. As stressed in Dixit et al. (2000), "*the ruling individuals must perceive an appreciable chance that their power will come to an end. They must be patient, so that future benefits can compensate for immediate concessions. And they must foresee a possibility of regaining power once it is lost.*" Moreover, parties must be "*policy-motivated*" (Bernhardt et al., 2009) so that they are attracted by policy goals as well as by the desire to win. Once elected, officeholders must be able to adopt whatever policy they consider most opportune, regardless of pressure from the opposition, lobbies or even their own party. In other words, "*the officeholder in these models becomes an elected dictator, governing all by himself with no legislature and no political parties*" (Usher, 2005). Last but not least, voters must "*hold politicians responsible for the policy choices that they make*" (Padovano & Petrarca, 2014). They also are "*rational, forward-looking, and informed*" so that voting decisions are based upon the rational expectations of the policies that the parties would follow if elected (Alesina, 1988).

Given the underlying assumptions, the responsible party government hypothesis seems unlikely. Yet, against some expectations, the empirical evidence suggests in most cases that officeholders differentiate their policies and fulfill most of their pledges once elected. Understanding how political markets achieve such outcomes is at the heart of this essay.

The outline of the review is as follows. Sections 2 and 3 focus on the tension that exits between the short run, where elections are held and incentives to be moderate are high, and the long run, where party credibility and policy differentiation matter most. Section 4 addresses the role of misinformation in electoral competition. Section 5

links the theory to the empirical literature. Section 6 concludes, and Sect. 7 extends the discussion.

2 Political Strategies in the Short Run

Many models of electoral competition such as the Downsian model or the citizen-candidate model are atemporal in the sense that they do not explicitly refer to past or future and, therefore, do not address the question of party credibility over time. Such a setting can be viewed as a model of the world where candidates are impatient and care mostly about short run incentives (see also Duggan & Martinelli, 2017). In those studies, the debate circles around one question: whether candidates compromise by announcing platforms that are desirable for a majority of voters, or formulate policies that are in line with their ideological views (for reviews, see Persson & Tabellini, 2000; Mueller, 2003; Padovano, 2013; De Donder & Gallego, 2017). As we will see below, the answer mainly depends on whether the running candidates are office-motivated (Sect. 2.1) or ideological (Sect. 2.2), and whether running for office is costly (Sect. 2.3).

2.1 Office-Motivated Politicians

The traditional spatial theory of voting explains the positioning of existing candidates, or parties, on a left–right scale. A well-known example in the field originates in the work of Hotelling (1929), Black (1948) and Downs (1957). The model, universally known as *Downsian model* or *median voter model*, considers two office-motivated candidates who care only about winning the election. Both candidates locate their platform on the left–right scale and keep their promises once elected. All citizens vote for the platform closest to their most-preferred position. As a result, an electoral equilibrium exists at the median voter's ideal point: candidates are incentivized to adopt moderate platforms if they want to attract a majority of votes (see Meltzer & Richard, 1981, for an extension to redistributive policies).

Platform convergence can also be achieved under a probabilistic setting when voters' decisions are erratic and based on probability distributions over positions. Since the candidates are assumed to be office-motivated, and face a symmetric optimization problem, they will announce the same platform if an equilibrium exists (Hinich, 1977, 1978; Lindbeck & Weibull, 1987, 1993; Coughlin, 1992; Ordeshook, 1986). Both the deterministic and probabilistic versions of the Downsian model thus conclude to the centripetal tendency of two-party competition.

Platform convergence is sometimes viewed as a desirable outcome from a purely normative view. For instance, under a probabilistic setting, platforms are shown to reach the *utilitarian point* (Hinich, 1977, 1978; Lindbeck & Weibull, 1987, 1993) or the *Nash point* (Coughlin & Nitzan, 1981), as if the candidates were maximizing

a social welfare function. This view, however, contradicts a large literature, which suggests instead that political parties should compete for citizen support on deliverable and contrasted positions (e.g., Bernhardt et al., 2009). Abstention is also likely to reach dramatic levels if parties have identical platforms (Downs, 1957).

2.2 Ideological Candidates

The second set of models views the politicians as "ideological", that is, with specific policy-oriented preferences. Simply put, candidates act to maximize their satisfaction which is a function of the observed political outcome. For instance, a left-wing party may advocate higher spending levels than a right-wing party for the simple reason that its members are prone to redistribution. In this view, political parties incur a utility loss when converging to moderate platforms.

The ideological nature of the candidates, however, is not a sufficient condition for observing platform divergence. Calvert (1985) shows that policy convergence is still achieved under perfect information. The running candidates will find it optimal to secure elections with moderate platforms as they are better off with a policy compromise than with the opponents' ideal point. Additional conditions are actually required for observing policy differentiation.

First, platforms can diverge under the assumption that candidates have imperfect information about the distribution of voters (Calvert, 1985; Hansson & Stuart, 1984; Roemer, 1994; Wittman, 1983, 1990). It may be irrelevant for the running candidates to move towards the center of the political spectrum if that move does not secure victory and, in the meantime, yields policy outcomes that are far from their ideological stance.

Second, divergence can be observed under ideological polarization, when voters at the extremes of the political spectrum refrain from voting because they feel that their interests are not represented by any of the running parties. Each party would gain more votes by holding their ideological positions rather than proposing moderate policies (Downs, 1957). This assumption has been central to many studies (Plane & Gershtenson, 2004).

Third, a set of studies have incorporated valence judgments in their model. Following Stokes' (1963) criticism of the Downsian model, voters are assumed to care about candidates' characteristics in addition to policy, such as charisma or the ability to deliver services to constituents (e.g., Ansolabehere & Snyder, 2000; Aragones & Palfrey, 2002; Gouret et al., 2011; Groseclose, 2001; Kartik & McAfee, 2007; Schofield, 2003). If one candidate has a particular advantage on a valence issue, then victory is secured almost regardless of where the candidates are located in the policy space (Ansolabehere & Snyder, 2000). Moreover, candidates who change their position toward the center could signal their type and be perceived as a lacking character (Kartik & McAfee, 2007).

Last, platform divergence can be conditioned by the rules and features of the electoral system (Alesina & Rosenthal, 1996, 2000; Gerber & Ortuno-Ortin, 1998;

Myerson, 1993). For instance, Gerber and Ortuno-Ortin (1998) examines a situation in which the implemented policy is a weighted combination of the proposed policies. In equilibrium, parties have incentives to make policy announcements that are more extreme than their own position, knowing that a compromise will be reached afterwards.

2.3 Endogenous Entry of Candidates

A limitation of the previous literature is that the number of running candidates, or parties, is exogenously specified. Osborne and Slivinski (1996) and Besley and Coate (1997) relax this assumption in their citizen-candidate models. Both studies consider a population of citizens who have preferences of their own. Each of them can choose to enter the electoral competition at some commonly known entry cost. In return, they receive a benefit if they take office. Because each ideal point is public information, the candidate who wins the election is assumed to implement his/her preferred policy. The argument is taken from Alesina (1988) in that "*promising anything else is not credible*" (Besley & Coate, 1997, p. 88).

Several specificities are worth being mentioned (Usher, 2005). First, while Osborne and Slivinski (1996) focus exclusively on a one-dimensional model, Besley and Coate (1997) are able to handle multidimensional issues and policy spaces. Second, while citizens vote strategically in Besley and Coate (1997), they vote sincerely in Osborne and Slivinski (1996) in the sense that they always vote for the candidate whose policy preference is closest to theirs. Third, Osborne and Slivinski (1996) consider a fixed cost of running for office and a fixed reward to the successful candidate. As a result, the equilibrium number of candidates depends negatively on the cost of running and positively on the benefits of winning. In contrast, in Besley and Coate (1997), the potential benefit from running is endogenous and derived strictly from policy outcomes.

Despite the differences mentioned above, the two models share similarities. They both depart from the previous literature in that they explain why candidates or parties have preferences of their own. They also provide an explanation for the variation in the number of candidates running for office and the differentiation of the competing platforms.

3 Political Strategies in the Long Run

This second set of studies examines with more attention how political candidates are incentivized to commit to their platform. Politicians are assumed to discount future payoffs—the more they care about their reputation, the more likely they are to keep their electoral promises. Specifically, their *reputation* relates positively to the extent to which policy positions are stable over time. The role of voters is particularly salient.

They do not only have policy preferences; they also evaluate each candidate from their past and/or expected actions, and vote accordingly. In the long run, the equilibrium strategies are thus shown to depend on the type of rationality involved in voting behavior (see Reed & Cho, 1998; Ashworth, 2012; Jan, 2015; Duggan & Martinelli, 2017, for reviews). Voting can be retrospective (Sect. 3.1), forward-looking (Sect. 3.2) or prospective (Sect. 3.3).

3.1 When Voting is Retrospective

Voters are said to be backward-looking when they adopt retrospective strategies in order to punish political parties. Studies examining this type of rationality include Key (1966), Fiorina (1981), Ferejohn (1986), Hibbs et al. (1981), Reed and Cho (1998), Duggan (2000), van Weelden (2013), among others. Voters are assumed to evaluate the government on the basis of its perceived past performance. Re-election then occurs only if the incumbent meets or exceeds a given performance standard. As argued in Aragonès et al. (2007), such a move from the voters cannot be credible in the long run. When faced with new elections, the platform of the incumbent could still look better than those of the opponents (see also Reed & Cho, 1998).

Retrospective voting has been the keystone of the traditional political business cycle. For Nordhaus (1975) and Tufte (1978), voters select candidates based on a retrospective evaluation of their performance. They are considered myopic and value only recent experiences. Politicians are assumed to be office-motivated and, knowing that voters can be easily fooled, engage in the stimulation of the economy in pre-election periods in order to get reelected. In other words, governments adopt expansionary fiscal policies before elections regardless of their ideology or economic factors. Here, policy convergence is achieved at the cost of unnecessary fluctuations.

3.2 When Voters Are Forward-Looking

Following the *rational expectations revolution* that occurred in the 1970's, several studies assume that voters do not make systematic mistakes in forecasting the future, at least on average. This is exemplified in Barro (1974) and the *Ricardian equivalence proposition* which assumes that consumers internalize the government's budget constraint when choosing their consumption levels. Public spending can thus be financed equivalently out of current taxes or current deficits. Both types of financing are equivalent because rational consumers will forecast the future taxes induced by the current deficits. The traditional political business cycle becomes inoperative. Yet, another outcome may be observed, where parties commit to their platforms and have different objectives concerning inflation and unemployment in accordance with the preferences of their core constituents (Alesina, 1987). In this view, left-wing parties

would be more inclined to stimulate aggregate demand than right-wing parties. Policies would thus differ depending on the elected party, the so-called "partisan cycle theory of macroeconomic policies".

More generally, Alesina (1988) and Alesina and Spear (1988) assume that the electorate is capable of forming rational expectations of the policies that would be followed by the candidates if elected. The results of partial or complete convergence are shown to be time inconsistent. During their last term in office, officeholders will always maximize their own satisfaction regardless of the announced platforms. Hence, voters will not believe any platform announcements that deviate from the candidates' ideological positions. In other words, candidates are "*identified with certain positions and cannot, without loss of credibility and trustworthiness, change these positions opportunistically*" (Holler & Skott, 2005). Holler and Skott (2005) note, however, that election agendas vary considerably from one election to the next, activating different sets of influences if voters are rationality bounded, which may in return explain alternations in power.

There could be a potential conflict between the short run interests of the individual policymakers, who are most likely to implement their favorite policy once in office, and those of the party, which in the long run has an incentive to announce moderate platforms in order to win elections (Alesina & Spear, 1988). According to Harrington (1992), the last-period effect is however reduced if incumbents wish their successors to be from the same political family so that future policy outcomes do not depart much from their ideal points. Reversely, whether policy convergence is observed, and promises are fulfilled, depends on the type of mechanism which a party uses to credibly establish moderate platforms (Alesina & Spear, 1988; Harrington, 1992).

More recently, Duggan and Fey (2006) investigate a repeated Downsian game where parties and voters anticipate the effects of their actions on future elections in a fully rational manner. They show that the median voter theorem holds if parties and voters are sufficiently impatient. Daley and Snowberg (2011) address the question of campaign financing. Politicians decide whether to exert effort toward implementing policy or toward raising funds. Voters cast their ballots based solely on the expected quality of candidates' future policies. Yet, since high-ability politicians can signal their type through raising more funds, the candidates may inefficiently allocate their efforts toward fund raising.

3.3 When Voting is Prospective

An alternative view of voting behavior involves the prospective outlook of voters which includes both aspects of backward-looking and forward-looking behaviors. A rational voter does not only prefer a government with better expected ability but also updates his/her belief through time based on past experiences. For instance, Cukierman and Meltzer (1986) and Rogoff (1987) address the question of the political business cycle, assuming that voters are able to monitor the government perfectly yet

with some lag. Because of this delay, welfare can be reduced under the assumption of office-motivated policymakers.

Several studies establish equilibria that are both consistent with retrospective and forward-looking behaviors. In Duggan (2000), incumbents win re-election if their most recent policy choice gives the median voter a payoff at least as high as he or she would expect from a challenger (see also Jan, 2015). Fishman and Klunover (2020) analyze equilibrium outcomes under uncertainty (e.g., possible threat to national security or a natural disaster) and show that incumbent adopts policies that are more aligned with voters' interests when the electorate is backward-looking. If the electorate is forward-looking, incumbents have incentives to ignore the extra information they have access to. In Shi and Svensson (2006), the size of electoral budget cycles is shown to depend on the share of voters that is informed about the amount of borrowing.

Several developments also show evidence of an incumbency advantage. Bernhardt and Ingberman (1985) argue that the incumbent represents a less risky lottery since the voters have observed the incumbent's actions. Banks and Duggan (2008) offer a model of repeated elections in which politicians determine policies in a multidimensional issue space. Because an untried challenger is inherently risky and voters are risk averse, an incumbent has a degree of leverage in choosing policies to achieve re-election.

4 Political Misinformation

Most voting theories take the uncertainty on the candidates' location as given. However, it may be rational for candidates to increase that uncertainty in order to avoid offending some constituents who hold opposite opinions. Downs (1957) argues on this matter that "*since both parties find it rational to be ambiguous, neither is forced by the other's clarity to take a more precise stand. […] True, their tendency towards obscurity is limited by the desire to attract voters to the polls, since citizens abstain if all parties seem identical or no party makes testable promises*". In this view, an optimal level of ambiguity could be targeted by the running candidates (Sect. 4.1). Interestingly, several authors have also considered the question of false promises. Here again, the extent to which voters are rational is essential in the formation of equilibria. Voters could be fooled by political promises and make their decisions based on false information: "*a party which perennially makes false promises can gain votes if it convinces voters to believe its lies*" (Downs, 1957). Yet, since lying is inherent in the political game, voters could also use the little information they have to approximate the true motivation of politicians (Sect. 4.2).

4.1 Strategic Ambiguity

Several studies have examined whether ambiguity is a relevant strategy during the election campaign phase. In most cases, *ambiguity* relates to the variance of the probability distributions over candidates' positions. The larger the variance, the less likely are the citizens to identify the true position. Zeckhauser (1969) and Shepsle (1972) argue for instance that candidates can be characterized as lotteries. Under that framework, a rational candidate who only cares about being in office will never choose to be ambiguous if voters are risk averse. Incentives to lie are minimized (Shepsle, 1972). Similarly, McKelvey (1980) examines the effects of the introduction of fixed amounts of ambiguity. Candidates have control over the position of their ambiguous strategies (location of their probability function) but not over the amount of ambiguity. It is shown that the introduction of ambiguity in a model where there was none before does not disrupt equilibria which existed in the original model.

Whether ambiguity is a relevant choice for the running candidates actually depends on several factors. Glazer (1990) shows that, if candidates are uncertain about the ideal position of the median voter, they could face the risk of stating an unpopular position. Hence, in equilibrium, the candidates may prefer to make their positions more ambiguous. In Aragones and Neeman (1994), candidates have a preference for ambiguity for the reason that a larger policy-set provides a wider margin for maneuvering in case of victory. Similarly, ambiguity can be a relevant strategy during the primary election phase since any precise commitment to a policy in the primary makes the candidates vulnerable to a vaguer challenger in the general election (Meirowitz, 2005). In Alesina and Cukierman (1990), the level of ambiguity is defined as the variance of the noise between the policy outcome observed by voters and the policy instrument chosen by politicians. Higher ambiguity enables the candidates to exploit the tradeoff between their ideology and the likelihood of reappointment (see also Enelow & Munger, 1993). Page (1976) and Rovny (2012) note that the candidate's best strategy is actually to avoid issues of a divisive sort, and place as nearly as possible no emphasis on them. Last, the voters themselves can develop a taste for ambiguity (Callander & Wilson, 2008).

Political ambiguity can have adverse effects as well. Laslier (2006) shows that the voters' aversion to ambiguity lowers the benefits from ambiguous platforms. Bräuninger and Giger (2018) argue that the choice of ambiguous policy platforms results from a tradeoff between electoral advantages and the necessity of preserving the support of party activists. Aragonès and Postlewaite (2002) note that the candidates can be restricted in the set of beliefs that they can induce in voters.

4.2 False Promises and Dishonesty in Politics

The question of false promises has also received attention. A set of studies assumes that voters can rationally infer from the political promises what each candidate would actually do if elected. For instance, in Banks (1990), the winning candidate's true policy position is revealed after the election. Announcing a position far from this position is costly. Provided that this cost is sufficiently high, candidates will diverge and be distinguishable, allowing the voters to select the right candidate.

Similarly, Haan (2004) argues that voters are not fooled by false promises. Politicians yet have an incentive to lie, since, in equilibrium, voters expect them to do so. Callander and Wilkie (2007) show that this result holds with heterogeneous lying costs. Elections are not necessarily dominated by the zero-cost types because voters are led to disbelieve candidates who promise too much. Garoupa (1999) considers a politician who derives a loss from being publicly exposed as a dishonest politician. As long as newspapers are deterred from defamation, their auditing role is effective in deterring political dishonesty.

In some cases, voters may hold incorrect beliefs. For instance, Schultz (1995) assumes that the running parties have an informational advantage over the voters. They know the true state of the economy and yet will not reveal this information to voters in order to sell their views. As a result, the winning policy does not match the current state of the economy. Heidhues and Lagerlöf (2003) also note that the running candidates have a strong incentive to follow popular beliefs (i.e., the voters' prior) instead of their own information, for the reason that they will have a hard time convincing the voters that their private information has a heavier weight.

Last, in accordance with the literature on political selection (see, e.g., Dal Bó & Finan, 2018, for a recent review), Caselli and Morelli (2004) and Messner and Polborn (2004) examine the self-selection of dishonest agents in politics. Dishonest citizens have a comparative advantage in running for office because they have a lower opportunity cost of choosing a life in politics and will moreover be able to steal more, and be able to extract greater rewards from public office.

5 Empirical Evidence

As we have seen, the theoretical literature offers several types of results depending on the underlying rationality of voters and candidates. In particular, politicians and political parties are facing a tradeoff between holding their ideology stances, in order to establish credibility in the long run, and converging to more consensual policies, with the risk of making false promises. Whether one equilibrium outperforms the other depends on whether the voters pose a sufficient and credible threat so that officeholders are incentivized to fulfill their pledges and to propose clear distinctive policies. This has given rise to five fundamental empirical questions. First, do we observe convergence of parties to the same policies (Sect. 5.1)? Second, do

parties keep their political promises once elected (Sect. 5.2)? Third, is there a link between economic performance and incumbents' popularity (Sect. 5.3)? Fourth, does party identification matter for voting choices (Sect. 5.4)? Fifth, are voting behaviors impacted by irrelevant factors? (Sect. 5.5). The next sections address these questions sequentially.

5.1 Is Convergence of Policies Observed?

A large strand of the empirical literature has been devoted to analyzing the influence of party ideology on public policies. The underlying question is whether parties converge or diverge in behavior.

As a matter of fact, several studies show that left-wing governments spend and tax more than their right-wing counterparts (see, e.g., Potrafke, 2018 for a review about the United States). Recently, that view has found support in France (Foucault et al., 2008; Le Maux et al., 2011), Italy (Padovano & Petrarca, 2014; Santolini, 2008), Swedish local governments (Folke, 2014; Pettersson-Lidblom, 2008), German municipalities (Freier & Odendahl, 2015), as well as OECD countries (Pickering & Rockey, 2011). Exploiting a dataset of 800 estimates from papers published between 1992 and 2018, Magkonis et al. (2021) offers a meta-analysis that confirms the significant link between ideology and government spending.

In light of this constant evidence, the question remains about what mechanisms drive those ideological effects. As argued in Pettersson-Lidbom (2008) and Le Maux et al. (2020), government ideology is endogenous and induces political equilibria that are dependent on the electorate's preferences and characteristics. In short, left-wing governments could spend more not only because of their ideological position but also because they are elected in jurisdictions where the demand for public goods and/or redistribution is higher. A higher ideological polarization of voters could also explain a larger divergence of outcomes (Clarke & Stewart, 1984).

To disentangle those effects, several studies have applied quasi-experimental techniques and either confirm the ideological effects of party control (Caughey et al., 2017; Freier & Odendahl, 2015; Hill & Jones, 2017; Pettersson-Lidbom, 2008), or have conclusions which are conditioned by the type of policy or instrument which is examined (Folke, 2014; Gerber & Hopkins, 2011; Le Maux et al., 2020). Those results also raise methodological issues. For instance, a shortcoming of studies using a regression discontinuity design is that they only focus on close elections (Le Maux et al., 2020; Potrafke, 2018). Moreover, despite the observed political differentiation, it is possible that the resulting policies have little effect on economic outcomes within the timeline between elections, as suggested for instance in Holbein and Dynes (2018).

5.2 Do Parties Keep Their Political Promises?

The question of whether political parties keep their promises has been intensively investigated in political science. Kennedy et al. (2021) offers a recent review of the literature. Most studies suggest that political parties tend to keep a majority of their campaign promises in a variety of contexts and countries. For instance, Pétry and Collette (2009) examine 18 journal articles and book chapters published in English and French that report quantitative measures of election promise fulfillment in North America and Europe. They find that parties fulfill 67% of their promises on average, with wide variation across time, countries, and regimes. Thomson (2011) study the fulfillment of over 20,000 pledges made in 57 election campaigns in 12 countries. Their findings suggest that government executives are highly likely to fulfill their pledges. The literature also notes that promise fulfillment is dependent on the extent of political fragmentation (e.g., Moury & Fernandes, 2018) and the state of the economy (e.g., Praprotnik, 2017).

Reversely, the evidence on the self-selection of dishonest agents into politics is scarce and mitigated. On the one hand, Fehrler et al. (2016) show in their experiment that dishonest people over-proportionally self-select into the political race when voters have no information about the resources that have been spent in the entry contest. Hanna and Wang (2017) implement a laboratory task with students in India. Participants who cheat are more likely to prefer public sector jobs. On the other hand, Olsen et al. (2019) investigate a dice game and show how individual-level behavioral dishonesty is very strongly negatively correlated with public service motivation. In their study about Sweden, Bó et al. (2017) find that politicians are on average significantly "smarter" and "better" leaders than the population they represent, even when controlling for family background. Using survey and experimental data covering village councils in rural West Bengal, Chaudhuri et al. (2020) find that inexperienced village council politicians are less dishonest and more pro-social than ordinary citizens.

5.3 Is There a Link Between Economic Performance and Incumbents' Popularity?

Several studies have been devoted to the empirical investigation of the relationship between economic performance and incumbents' popularity. In most cases, the literature shows support for the retrospective hypothesis.

For instance, Kramer (1971) focuses on the vote for the U.S. House of Representatives and shows that election outcomes are in substantial part responsive to objective changes occurring under the incumbent party. Similarly, using data from the United States, Frey and Schneider (1978) find that economic events significantly influence presidential popularity. In contrast, Golden and Poterba (1980) examine the validity

of the political business cycle theory in the United States and find no important or statistically significant effects.

Hibbs et al. (1981) analyze the case of France and show that the performance of the real economy matters for future elections, and that the weights placed upon current and past performance in the formation of contemporaneous political judgments decline geometrically. Similar results are found in the response of political support for American presidents (Hibbs et al., 1982), for British (Hibbs, 1982a) and German (Hibbs, 1982b) governments.

Voting responses are not always symmetric. Hibbs (2000) suggests a bias favoring the incumbent party, which may be rationalized by risk aversion, since voters have more recent information about the party in power than about the opposition. Campbell et al. (2010) finds that retrospective evaluations are applied more strictly to incumbents seeking election than to successor candidates who are not incumbents. Using election data from three U.S. states, Bloom and Price (1975) show that economic downturns reduce the vote for the party of the incumbent President, but economic upturns have no corresponding effect. Similarly, in their analysis of the economic voting of 17,100 Danes, Nannestad and Paldam (1997) demonstrate that voters' reaction is about three times larger to a deterioration in the economy than to an improvement.

Evidence on whether voters rationally anticipate future outcomes is subject to controversy. For instance, using survey data about the US economy, Haller and Norpoth (1994) show that voter forecasts are prone to asymmetry. While good economic times tend to foster optimism in an adaptive fashion, bad times do not similarly feed pessimism. Using data from the European Union Consumer Confidence Surveys, Duch and Stevenson (2011) find that incumbent governments are less able to sway citizens to optimistic forecasts but opposition can make citizens feel inflation is actually worse.

5.4 Does Party Identification Matter for Voting Choices?

The literature in political science suggests that party identification is an important variable shaping voting behaviors (see, e.g., Evans & Chzhen, 2015; Lachat, 2015). As initially suggested in the traditional spatial voting theory, partisanship encourages voters to support a particular party. Several studies also suggest that the perceived policy performance and party competence (leadership, honesty, integrity, and trustworthiness) have a causal impact on voting choices, confirming the valence theory. As a matter of fact, both voting behaviors are shown to interact. This is exemplified in the partisan analysis of macroeconomic policies where the link between economic performance and incumbents' popularity appears to be class-related with different group responses to economic events (Hibbs et al., 1981, 1982; Hibbs, 1982a, 1982b).

Several studies have addressed the specific link between valence and party identification. For instance, using survey data from Ireland, Thomson (2011) show that citizens evaluate government performance more positively (resp. negatively) if they

identify with the running (resp. opposition) party. Using experimental survey data, Malhotra and Kuo (2008) provides a list of political officials involved in the preparation for and response to Hurricane Katrina and show that partisanship substantially affected whom citizens blamed for the effects of the hurricane. Using both survey and election data about Britain, Sanders et al. (2011) show that valence mechanisms outperform the spatial model in terms of strength of direct effects on voting choice. However, spatial effects still have sizeable indirect effects on the vote via their influence on valence judgments. Similarly, Gouret et al. (2011) show on French data that valence characteristics have a different impact on the utility of voters according to their position on the political spectrum. Using a crowdsourced survey experiment with German participants, Haselmayer et al. (2020) find that the perception of negative campaigning in a multi-party system is contingent on partisan preferences. Last, using data from a sample of around 2,000 British citizens, Johns and Kölln (2020) show that ideological moderation will boost a party's perceived competence. Less radical parties are seen as readier to compromise, more realistic about what can be achieved, and less prone to simplistic solutions.

5.5 Are Voting Behaviors Impacted by Irrelevant Factors?

Another stream of the literature examines whether voters are sufficiently informed about economic and political issues. Aidt (2000) provides several examples suggesting that voters are, on average, poorly informed about economic data such as the unemployment rate, inflation, growth in GDP and the level of public debt. They also know little about how the economic system works. Paldam and Nannestad (2000) show that voters are myopic and that people's expectations are largely static, i.e., voting is retrospective, but knowledge grows at election time. Using data on natural disasters, Healy and Malhotra (2009) find that voters reward the incumbent presidential party for delivering disaster relief spending, but not for investing in disaster preparedness spending. These inconsistencies distort the incentives of public officials, leading the government to underinvest in disaster preparedness. Last but not least, Jacoby (2000) shows that political rhetoric, and how political issues are framed, are significant sources of influence that politicians can use to alter public opinion. Similarly, Holler and Skott (2005) offer several examples suggesting that voters are rationally bounded.

A set of studies also suggest that election results are influenced by irrelevant events such as football games (Healy et al., 2010) or shark attacks (Achen & Bartels, 2002, 2016). Those results are subject to controversy (Fowler & Hall, 2018; Fowler & Montagnes, 2015; Gasper & Reeves, 2011). For instance, Gasper and Reeves (2011) find that electorates punish presidents and governors for severe weather damage. However, these effects are dwarfed by the response of attentive electorates to the actions of their officials. Fowler and Montagnes (2015) conclude that college football games do not meaningfully influence elections.

In addition, the personal characteristics of the running candidates may affect voting behaviors. Looking at the personalities of Belgian politicians, Joly et al. (2018) find correlations between personality traits and measures of political success. Todorov et al. (2005) show that inferences of competence based solely on facial appearance predict the outcomes of U.S. congressional elections better than chance. This creates incentives for the political parties to account for those considerations when selecting their running candidates. Based on estimates of the facial competence of 972 congressional candidates, Atkinson et al. (2009) find that in more competitive races the out-party tends to run candidates with higher quality faces.

Last, voting choices could be influenced by ideology-oriented media (DellaVigna & Kaplan, 2007; Durante & Knight, 2012; Sabatini, 2012) or biased by the usage of social network sites (Lee, 2020). Bessi and Ferrara (2016) and Bovet and Makse (2019) suggest that misinformation campaigns and websites spreading false news can alter public opinion and endanger the integrity of elections.

6 Conclusion

We have seen that the theory deals with different types of political equilibria, and that those equilibria are conditioned by the rationality of voters and running candidates. On the one hand, there is the interpretation that citizens care only about the current election, and vote for the candidate or party whose positions are closest to their own. If the pros outweigh the cons, a rational candidate will have an incentive to announce moderate policies to get elected. This might be referred to as the "Downsian conception" of the political market. The chances that campaign pledges are being fulfilled are however thin, since it is rational for officeholders to implement a policy that is more consistent with their ideological views. On the other hand, at the extreme, a set of models postulates "à la Alesina" that voters are able to build up their expectations as if they were fully informed of all political consequences. Because ideological candidates cannot credibly commit to moderate platforms, the only time-consistent electoral equilibrium is one in which divergence is observed, and promises are fulfilled. In this view, the *responsible party government hypothesis* does hold. Between these two extreme conceptions, i.e., opportunistic (purely office-motivated) parties and responsible (policy-motivated) parties, a set of studies consider the possibility for the candidates to be manipulative, dishonest or ambiguous on purpose, and the voters to be partially informed. In those cases, equilibria could be, but not necessarily, altered. Falsehoods may still be neutralized by rational voter skepticism.

Which of these three conceptions best explains the data? This might be a surprise to some readers but the empirical literature seems to support, *at least to some extent*, the responsible party government hypothesis. Political candidates are in most cases shown to hold to their political promises once elected; and left-wing governments are generally found to spend more than their right-wing counterparts. One possible reason that is emphasized in this essay is that while the voters are partially informed

(Aidt, 2000) and rationally bounded (Holler & Skott, 2005), they also care about both the ideology and performance of the running candidates. Hence, they could pose a sufficient and credible threat so that candidates are incentivized to propose clear distinctive platforms and to fulfill their pledges.

From this perspective, how can we explain the political distrust that is observed around the world? While the literature does not seem to reach a consensus, several explanations are identified. The first set of causes could be inherent to any representative structure. As established in Arrow's (1951) *impossibility theorem*, the political representation of multiple interests is in essence conflictual. The fact that there is no ideal voting system for making collective choices under heterogeneous preferences automatically generates political disaffection. A somewhat similar explanation lies in the *paradox of voting* (Downs, 1957). While voters can derive utility from the act of voting itself, they still might feel that voting is useless given the small probabilities of affecting electoral outcomes (see Lyytikäinen & Tukiainen, 2019, for recent empirical evidence). Citizens also have intrinsic values of decision rights in the sense that *freedom of choice* is important for a person's quality of life (Bartling et al., 2014). In this view, representative democracy would be rejected no matter what by those who which to have better control over their life.

Party identification also is a possible source of *misperception about government performance* (Malhotra & Kuo, 2008; Thomson, 2011). Voters who do not identify with a party may rate the performance of that party more negatively. In this perspective, high levels of ideological polarization are likely to be associated with highly negative views about the functioning of democracy. The role of other sources of misperception, resulting from heterogeneous information resources or personal experience (Thomson, 2011), ideology-oriented media (DellaVigna & Kaplan, 2007; Durante & Knight, 2012; Sabatini, 2012), social network sites (Lee, 2020) and negative campaigning (Haselmayer et al., 2020) are worth being mentioned.

Last but not least, false promises have *negative psychological consequences* (Sencer, 1991). As argued in Harrington (1992), it could be well that voters become skeptical as to the credibility of the whole party system, even if only one single party reneges. A similar argument lies in Tversky and Kahneman (1973) who suggest that individuals use heuristics and can assign higher subjective probabilities to extreme events. This could be particularly true in politics, where scandals and lies are intensively covered by media. Asymmetry in behavior is expected, as suggested in Naurin et al. (2019) who highlight a tendency for governments to be penalized for unfulfilled pledges more than they are rewarded for fulfilled pledges. As a result, politicians are likely to be more sensitive to real or potential losses than they are to gains, inducing agenda limitation, scapegoating, or even defection (Weaver, 1986).

To sum up, the conclusion of this essay is in line with Wittman's (1989) optimistic view that "*democratic political markets are organized to promote wealth-maximizing outcomes, that these markets are highly competitive, and that political entrepreneurs are rewarded for efficient behavior*". Political distrust can yet prevail for the simple reason that various factors affect how democracy is perceived: the political representation of multiple interests is in essence conflictual; evaluations of government performance are possibly biased; and citizens may overreact to pledge unfulfillment.

Which of these explanations best fits the data and does the resulting negativity bias affect policy outcomes surely offer a set of challenging questions for future research.

7 Discussion

This essay offers an up-to-date picture of the state of the art about electoral competition, with the aim of linking the competing theories with their empirical investigation. It provides a collection of the most relevant and significant publications. As stressed in the introduction, however, the review is not exhaustive, and several important considerations have been eluded. Some of them are discussed below.

First, most of the literature mentioned in this essay belongs to the Public Choice/Political Economy literature, PCPE hereafter, which can be defined as "*the economic study of non-market decision-making, or simply application of economics to political science*." (Mueller, 1989). The subject matter is relatively the same as that of political science, yet the methodology differs in that it relies more prominently on "*rational choice theory*" (Kurrild-Klitgaard, 2018). In the tradition of standard (neoclassical) microeconomic analysis, theories in the field are based on several mathematical assumptions, e.g., about the polity and the agents' objectives (see Winer & Ferris, 2022, who underlie the differences between an economic model of a market and a political economy model of public policy). Each element is shown to play a role, inducing multiple equilibria and propositions. The literature thus strongly reflects the discipline from which it comes, that is *economics*; and is therefore subject to the same criticisms.

Second, given their multiplicity and specificities, theories in the field require empirical testing and validation. This can be challenging: data are not always available; measurement errors are possible; confounding factors are in play. Challenges like these are inherent to any social science research, but they often put the validation procedure of PCPE theories in difficulties. With the increased availability of political data, however, and the adoption of new quantitative techniques, empirical studies in the field have been booming. A rising number of PCPE studies now rely on counterfactual methods to prove causation, improving the "scientific" stature of the discipline.

Third, the PCPE literature is vast and varied. The reader is invited to look at Mueller's (2003) major textbook for a well-established description of this literature. Many studies in the field are devoted to the analysis of government failures, their determinant and their prevalence. The basic tenet is that the agents evolving in the public sector are not different from those operating in the private sector. Governments can fail because of the self-serving behavior of politicians and bureaucrats; for instance, because of corruption and rent-seeking behaviors. More specifically, the discipline raises questions about whether government failures are inherently more or less severe than market failures. The question of the allocative efficiency and equity of public policies is also central in public economics, another branch of economics.

Moreover, a set of PCPE models address the economic determinants of electoral preferences. Most of these works are rooted in the tradition of Samuelson's (1954) theory of public goods. Heterogeneity in taste, income and tax base are shown to induce heterogeneous preferences; each voter is characterized by a demand for private and public goods. Whether a voter benefits from higher public good provision (and higher tax rates) depends on how he or she values the public goods compared to the private goods. Additionally, electoral preferences are shaped by partisanship, altruism, inequality aversion, and other behavioral factors. A rising number of studies address the relative importance of these items (e.g., in the context of randomized controlled experiments). This individual heterogeneity further raises concerns about preference representation in democracies, and receives particular attention in social choice theory.

Relatedly, several studies examine the impact of electoral rules on political market structures. The outlines are the following. The more proportional is an electoral rule, the higher the chances that a small party gets seats, which reduces the need of forming pre-electoral coalitions (Golder, 2005, 2006; Hortala et al., 2021) and lowers the likelihood of platform convergence (Matakos et al., 2016). Thus, in line with Duverger's (1964) rules, proportional voting systems favor party fragmentation (i.e., oligopolistic market structures, although this term is rarely employed in the PCPE literature), while majoritarian systems tend to favor a two-party system. Proportional systems are furthermore shown to reduce the degree of disproportionality between the distribution of seats in a legislature and the distribution of votes (Lundell, 2012); and to increase the quality of political representation (Stadelmann et al., 2014).

Last but not least, this essay examines two particular conditions for a good functioning of representative democracies—whether political parties have divergent positions and whether they fulfill their promises—but disregards several important qualitative dimensions, such as *freedom*, the *rule of law*, *responsiveness*, *equality*, and *participation* (Diamond & Morlino, 2005). The reason is methodological. Most of PCPE studies about electoral competition focuses on liberal democracies, i.e., assume an exogenous democratic rule as if political institutions were stable and competitive. As stressed in Winer and Ferris (2022), "*the use of the adjective "liberal" acknowledges the written laws and unwritten conventions that limit the extent to which any citizen may be forced to accept the consequences of government actions. These constraints on governing power significantly shape the political process.*" Note that several authors relax these assumptions, and consider a longer time horizon over which regimes evolve. Voters have preferences over regimes in accordance with their ideology, and account for the economic and social consequences of these regimes. The role of inequality and distributive conflict are then shown to play a role in transitions to and from the democratic rule (see, e.g., Acemoglu & Robinson, 2006).

While the analogy with the traditional analysis of private markets is straightforward, and if the concepts (supply and demand, the concept of a representative agent) and methodologies (economics of asymmetric information, intertemporal choices) are somewhat similar, several important demarcations exist between *the consumer and producer theory* and PCPE. The essential differences lie in that an election is a collective process in which the voters compromise; various electoral rules exist; and

several factors influence the agents' decisions, other than profit and consumption maximization. This makes the discipline all the more singular and challenging.

References

Acemoglu, D., & Robinson, J. A. (2006). *Economic origins of dictatorship and democracy*. Cambridge University Press.

Achen, C. H., & Bartels, L. M. (2002). *Blind retrospection: Electoral responses to drought, flu, and shark attacks*. Presented at the annual meeting of the American Political Science Association.

Achen, C. H., & Bartels, L. M. (2016). *Democracy for realists: Why elections do not produce responsive government*. Princeton University Press.

Aidt, T. S. (2000). Economic voting and information. *Electoral Studies, 19*, 349–362.

Alesina, A. (1987). Macroeconomic policy in a two-party system as a repeated game. *Quarterly Journal of Economics, 102*, 651–678.

Alesina, A. (1988). Credibility and policy convergence in a two-party system with rational voters. *American Economic Review, 78*, 796–805.

Alesina, A., & Cukierman, A. (1990). The politics of ambiguity. *Quarterly Journal of Economics, 105*, 829–850.

Alesina, A., & Rosenthal, H. (1996). A theory of divided government. *Econometrica, 64*, 1311–1341.

Alesina, A., & Rosenthal, H. (2000). Polarized platforms and moderate policies with checks and balances. *Journal of Public Economics, 75*, 1–20.

Alesina, A., & Spear, S. (1988). An overlapping generations model of electoral competition. *Journal of Public Economics, 37*, 359–379.

Ansolabehere, S., & Snyder, J. M. (2000). Valence politics and equilibrium in spatial election models. *Public Choice, 103*, 327–336.

Aragones, E., & Neeman, Z. (1994). Strategic ambiguity in electoral competition. *Journal of Theoretical Politics, 12*, 183–204.

Aragones, E., & Palfrey, T. (2002). Mixed equilibrium in a Downsian model with a favored candidate. *Journal of Economic Theory, 103*, 131–161.

Aragonès, E., Palfrey, T., & Postlewaite, A. (2007). Political reputations and campaign promises. *Journal of the European Economic Association, 5*, 846–884.

Aragonès, E., & Postlewaite, A. (2002). Ambiguity in election games. *Review of Economic Design, 7*, 233–255.

Arrow, K. (1951). *Social choice and individual values* (2nd ed.). Yale University Press.

Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science, 15*, 183–201.

Atkinson, M., Enos, R., & Hill, S. (2009). Candidate faces and election outcomes: Is the face-vote correlation caused by candidate selection? *Quarterly Journal of Political Science, 4*, 229–249.

Bó, E., Finan, F., Folke, O., Persson, T., & Rickne, J. (2017). Who becomes a politician? *Quarterly Journal of Economics, 132*, 1877–1914.

Banks, J. S. (1990). A model of electoral competition with incomplete information. *Journal of Economic Theory, 50*, 309–325.

Banks, J. S., & Duggan, J. (2008). A dynamic model of democratic elections in multidimensional policy spaces. *Quarterly Journal of Political Science, 3*, 269–299.

Barro, R. J. (1974). Are government bonds net wealth? *Journal of Political Economy, 82*, 1095–1117.

Bartling, B., Fehr, E., & Herz, H. (2014). The intrinsic value of decision rights. *Econometrica, 82*, 2005–2039.

Bernhardt, D., Duggan, J., & Squintani, F. (2009). The case for responsible parties. *American Political Science Review, 103*, 570–587.

Bernhardt, D. M., & Ingberman, D. E. (1985). Candidate reputations and the incumbency effect. *Journal of Public Economics, 27*, 47–67.
Besley, T., & Coate, S. (1997). An economic model of representative democracy. *Quarterly Journal of Economics, 112*, 85–114.
Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, 21.
Black, D. (1948). On the rationale of group decision making. *Journal of Political Economy, 56*, 23–34.
Bloom, H. S., & Price, H. D. (1975). Voter response to short-run economic conditions: The asymmetric effect of prosperity and recession. *American Political Science Review, 69*, 1240–1254.
Bovet, A., & Makse, H. (2019). Influence of fake news in twitter during the 2016 US Presidential election. *Nature Communications*, 10.
Bräuninger, T., & Giger, N. (2018). Strategic ambiguity of party positions in multi-party competition. *Political Science Research and Methods, 6*, 527–548.
Buchanan, J., & Tullock, G. (1962). *The calculus of consent: Logical foundations of constitutional democracy*. Liberty Fund Inc.
Callander, S., & Wilkie, S. (2007). Lies, damned lies, and political campaigns. *Games and Economic Behavior, 60*, 262–286.
Callander, S., & Wilson, C. H. (2008). Context-dependent voting and political ambiguity. *Journal of Public Economics, 92*, 565–581.
Calvert, R. (1985). Robustness of the multidimensional voting model: Candidates' motivations, uncertainty and divergence. *American Journal of Political Science, 29*, 69–95.
Campbell, J. E., Dettrey, B. J., & Yin, H. (2010). The theory of conditional retrospective voting: Does the presidential record matter less in open-seat elections? *Journal of Politics, 72*, 1083–1095.
Caselli, F., & Morelli, M. (2004). Bad politicians. *Journal of Public Economics, 88*, 759–782.
Caughey, D., Xu, Y., & Warshaw, C. (2017). Incremental democracy: The policy effects of partisan control of state government. *The Journal of Politics, 79*, 1342–1358.
Chaudhuri, A., Iversen, V., Jensenius, F. R., & Maitra, P. (2020). Selecting the best of us? Politician quality in village councils in west Bengal, India. CESifo Working Paper Series, 8597.
Clarke, H. D., & Stewart, M. C. (1984). Dealignment of degree: Partisan change in Britain, 1974–83. *Journal of Politics, 46*, 689–718.
Committee on Political Parties (1950). Toward a more responsible two-party system: A report of the committee on political parties. *American Political Science Review*, 44(Part 2, Suppl.).
Coughlin, P. (1992). *Probabilistic voting theory*. Cambridge University Press.
Coughlin, P., & Nitzan, S. (1981). Electoral outcomes with probabilistic voting and Nash social welfare maxima. *Journal of Public Economics, 15*, 113–122.
Cronin, J. (2019). Responsible parties, responsible government and the present crisis. In F. McCall Rosenbluth, & I. Shapiro (Eds.), *Responsible parties: Saving democracy from itself* (pp. 1–8). New Haven, CT: Yale University Press.
Cukierman, A., & Meltzer, A. H. (1986). A positive theory of discretionary policy, the cost of democratic government and the benefits of a constitution. *Economic Inquiry, 24*, 367–388.
Dal Bó, E., & Finan, F. (2018). *Progress and perspectives in the study of political selection. Working Paper 24783.* National Bureau of Economic Research.
Daley, B., & Snowberg, E. (2011). Even if it is not bribery: The case for campaign finance reform. *Journal of Law, Economics, & Organization, 27*, 324–349.
De Donder, P., & Gallego, M. (2017). Electoral Competition and Party Positioning. *TSE Working Papers 17–760, Toulouse School of Economics (TSE).*
DellaVigna, S., & Kaplan, E. (2007). The fox news effect: Media bias and voting. *Quarterly Journal of Economics, 122*, 1187–1234.
Diamond, L., & Morlino, L. (2005). *Assessing the quality of democracy*. Johns Hopkins University Press.
Dixit, A., Grossman, G. M., & Gul, F. (2000). The dynamics of political compromise. *Journal of Political Economy, 108*, 531–568.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper and Row.
Duch, R., & Stevenson, R. (2011). Context and economic expectations: When do voters get it right? *British Journal of Political Science*, 41.
Duggan, J. (2000). Repeated elections with asymmetric information. *Economics & Politics, 12*, 109–135.
Duggan, J., & Fey, M. (2006). Repeated Downsian electoral competition. *International Journal of Game Theory, 35*, 39–69.
Duggan, J., & Martinelli, C. (2017). The political economy of dynamic elections: Accountability, commitment, and responsiveness. *Journal of Economic Literature, 55*, 916–984.
Durante, R., & Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from Berlusconi's Italy. *Journal of the European Economic Association, 10*, 451–481.
Duverger, M. (1964). *Political parties: their organization and activity in the modern state*. London, Methuen: Internet Archive.
Enelow, J. M., & Munger, M. C. (1993). The elements of candidate reputation: The effect of record and credibility on optimal spatial location. *Public Choice, 77*, 757–772.
Evans, G., & Chzhen, K. (2015). Re-evaluating the valence model of political choice. *Political Science Research and Methods, 4*, 199–220.
Fehrler, S., Fischbacher, U., & Schneider, M. T. (2016). Who runs? Honesty and self-selection into politics. *IZA Discussion Papers 10258*. Institute of Labor Economics (IZA).
Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice, 50*, 5–25.
Fiorina, M. (1981). *Retrospective voting in American elections*. Yale University Press.
Fishman, A., & Klunover, D. (2020). To Act or not to Act? *Political competition in the presence of a threat*. Papers 2010.03464, arXiv.org.
Folke, O. (2014). Shades of brown and green: Party effects in proportional election systems. *Journal of the European Economic Association, 12*, 1361–1395.
Foucault, M., Madies, T., & Paty, S. (2008). Public spending interactions and local politics. Empirical evidence from French municipalities. *Public Choice, 137*, 57–80.
Fowler, A., & Hall, A. B. (2018). Do shark attacks influence presidential elections? Reassessing a prominent finding on voter competence. *Journal of Politics, 80*, 1423–1437.
Fowler, A., & Montagnes, B. P. (2015). College football, elections, and false-positive results in observational research. *Proceedings of the National Academy of Sciences of the United States of America, 112*, 13800–13804.
Freier, R., & Odendahl, C. (2015). Do parties matter? Estimating the effect of political power in multi-party systems. *European Economic Review, 80*, 310–328.
Frey, B. S., & Schneider, F. (1978). An empirical study of politico-economic interaction in the United States. *Review of Economics and Statistics, 60*, 174–183.
Garoupa, N. (1999). The economics of political dishonesty and defamation. *International Review of Law and Economics, 19*, 167–180.
Gasper, J. T., & Reeves, A. (2011). Make it rain? Retrospection and the attentive electorate in the context of natural disasters. *American Journal of Political Science, 55*, 340–355.
Gerber, A., & Ortuno-Ortin, I. (1998). Political compromise and endogenous formation of coalitions. *Social Choice and Welfare, 15*, 445–454.
Gerber, E. R., & Hopkins, D. J. (2011). When mayors matter: Estimating the impact of mayoral partisanship on city policy. *American Journal of Political Science, 55*, 326–339.
Glazer, A. (1990). The strategy of candidate ambiguity. *American Political Science Review, 84*, 237–241.
Golder, S. N. (2005). Pre-electoral coalitions in comparative perspective: A test of existing hypotheses. *Electoral Studies, 24*, 643–663.
Golder, S. N. (2006). Pre-electoral coalition formation in parliamentary democracies. *British Journal of Political Science, 36*, 193–212.
Golden, D. G., & Poterba, J. M. (1980). The price of popularity: The political business cycle reexamined. *American Journal of Political Science, 24*, 696–714.

Gouret, F., Hollard, G., & Rossignol, S. (2011). An empirical analysis of valence in electoral competition. *Social Choice and Welfare, 37*, 309–340.
Groseclose, T. (2001). A model of candidate location when one candidate has a valence advantage. *American Journal of Political Science, 45*, 862–886.
Haan, M. (2004). Promising politicians, rational voters, and election outcomes. *Spanish Economic Review, 6*, 227–241.
Haller, H. B., & Norpoth, H. (1994). Let the good times roll: The economic expectations of U.S. voters. *American Journal of Political Science, 38*, 625–650.
Hanna, R., & Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy, 9*, 262–290.
Hansson, I., & Stuart, C. (1984). Voting competitions with interested politicians: Platforms do not converge to the preference of the median voter. *Public Choice, 44*, 431–442.
Harrington, J. (1992). The role of party reputation in the formation of policy. *Journal of Public Economics, 49*, 107–121.
Haselmayer, M., Hirsch, L., & Jenny, M. (2020). Love is blind. Partisanship and perception of negative campaign messages in a multiparty system. *Political Research Exchange*, 2, 1806002.
Healy, A., & Malhotra, N. (2009). Myopic voters and natural disaster policy. *American Political Science Review, 103*, 387–406.
Healy, A. J., Malhotra, N., & Mo, C. H. (2010). Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences, 107*, 12804–12809.
Heidhues, P., & Lagerlöf, J. (2003). Hiding information in electoral competition. *Games and Economic Behavior, 42*, 48–74.
Hibbs, D. (1982a). Economic outcomes and political support for British governments among occupational classes: A dynamic analysis. *American Political Science Review, 76*, 259–279.
Hibbs, D. (1982b). On the demand for economic outcomes: Macroeconomic performance and mass political support in the United States, Great Britain, and Germany. *Journal of Politics, 44*, 426–462.
Hibbs, D., Vasilatos, N., & Hibbs, D., Jr. (1981). Economics and politics in France: Economic performance and mass political support for Presidents Pompidou and Giscard d'Estaing. *European Journal of Political Research, 9*, 133–145.
Hibbs, D. A. (2000). Bread and peace voting in U.S. Presidential Elections. *Public Choice, 104*, 149–180.
Hibbs, D. A., Rivers, R. D., & Vasilatos, N. (1982). The dynamics of political support for American presidents among occupational and partisan groups. *American Journal of Political Science, 26*, 312–332.
Hill, A. J., & Jones, D. B. (2017). Does partisan affiliation impact the distribution of spending? Evidence from state governments' expenditures on education. *Journal of Economic Behavior & Organization, 143*, 58–77.
Hinich, M. J. (1977). Equilibrium in spatial voting: The median voter result is an artifact. *Journal of Economic Theory, 16*, 208–219.
Hinich, M. J. (1978). Some evidence on non-voting models in the spatial theory of electoral competition. *Public Choice, 33*, 83–102.
Holbein, J. B., & Dynes, A. (2018). *Noisy retrospection: The effect of party control on policy outcomes*. Working paper, Department of Political Science, Brigham Young University.
Holler, M. J., & Skott, P. (2005). Election campaigns, agenda setting and electoral outcomes. *Public Choice, 125*, 215–228.
Hortala-Vallve, R., Meriläinen, J., & Tukiainen, J. (2021). *Pre-electoral coalitions: Insights into the boundaries of political parties*. Technical report, APSA Preprints.
Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*, 41–57.
Jan, K. (2015). Lexicographic voting: Holding parties accountable in the presence of Downsian competition. *B.E. Journal of Economic Analysis & Policy, 15*, 1867–1892.
Jacoby, W. G. (2000). Issue framing and public opinion on government spending. *American Journal of Political Science*, 44(4), 750–767.

Johns, R., & Kölln, A.-K. (2020). Moderation and competence: How a party's ideological position shapes its valence reputation. *American Journal of Political Science, 64*, 649–663.
Joly, J., Soroka, S., & Loewen, P. (2018). Nice guys finish last: Personality and political success. *Acta Politica*, 54.
Kartik, N., & McAfee, R. P. (2007). Signaling character in electoral competition. *American Economic Review, 97*, 852–870.
Kennedy, J., Alcantara, C., & Armstrong, D. A., II. (2021). Do governments keep their promises? An analysis of speeches from the throne. *Governance, 34*, 917–934.
Key, V. (1966). *The responsible electorate: Rationality in presidential voting, 1936–1960*. Cambridge, MA.
Kramer, G. H. (1971). Short-term fluctuations in U.S. voting behavior, 1896–1964. *American Political Science Review, 65*, 131–143.
Kurrild-Klitgaard, P. (2018). Public choice and political science: A view from Europe. *Public Choice, 175*, 245–257.
Lachat, R. (2015). The role of party identification in spatial models of voting choice. *Political Science Research and Methods, 3*, 641–658.
Laslier, J.-F. (2006). Ambiguity in electoral competition. *Economics of Governance, 7*, 195–210.
Le Maux, B., Dostálová, K., & Padovano, F. (2020). Ideology or voters? A quasi-experimental test of why left-wing governments spend more. *Public Choice, 182*, 17–48.
Le Maux, B., Rocaboy, Y., & Goodspeed, T. (2011). Political fragmentation, party ideology and public expenditures. *Public Choice, 147*, 43–67.
Lee, H. (2020). Voters' involvement, attitude, and confidence in the era of new media. *Palgrave Communications, 6*, 1–7.
Lindbeck, A., & Weibull, J. (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice, 52*, 273–297.
Lindbeck, A., & Weibull, J. (1993). A model of political equilibrium in a representative democracy. *Journal of Public Economics, 51*, 195–209.
Lundell, K. (2012). Disproportionality and party system fragmentation: Does assembly size matter? *Open Journal of Political Science*, 02.
Lyytikäinen, T., & Tukiainen, J. (2019). Are voters rational? *European Journal of Political Economy, 59*, 230–242.
Matakos, K., Troumpounis, O., & Xefteris, D. (2016). Electoral rule disproportionality and platform polarization. *American Journal of Political Science, 60*, 1026–1043.
Magkonis, G., Zekente, K.-M., & Logothetis, V. (2021). Does the left spend more? An econometric survey of partisan politics. *Oxford Bulletin of Economics and Statistics, 83*, 1077–1099.
Malhotra, N., & Kuo, A. G. (2008). Attributing blame: The public's response to hurricane Katrina. *Journal of Politics, 70*, 120–135.
McKelvey, R. D. (1980). Ambiguity in spatial models of policy formation. *Public Choice, 35*, 385–402.
Meirowitz, A. (2005). Informational party primaries and strategic ambiguity. *Journal of Theoretical Politics, 17*, 107–136.
Meltzer, A. H., & Richard, S. F. (1981). A rational theory of the size of government. *Journal of Political Economy, 89*, 914–927.
Messner, M., & Polborn, M. K. (2004). Paying politicians. *Journal of Public Economics, 88*, 2423–2445.
Moury, C., & Fernandes, J. M. (2018). Minority governments and pledge fulfilment: Evidence from Portugal. *Government and Opposition, 53*, 335–355.
Mueller, D. C. (1989). *Public choice II, a revised edition of public choice*. Cambridge University Press.
Mueller, D. C. (2003). *Public choice III*. Cambridge University Press.
Myerson, R. B. (1993). Incentives to cultivate favored minorities under alternative electoral systems. *American Political Science Review, 87*, 856–869.

Nannestad, P., & Paldam, M. (1997). The grievance asymmetry revisited: A micro study of economic voting in Denmark, 1986–1992. *European Journal of Political Economy, 13*, 81–99.
Naurin, E., Soroka, S., & Markwat, N. (2019). Asymmetric accountability: An experimental investigation of biases in evaluations of governments' election pledges. *Comparative Political Studies, 52*, 2207–2234.
Nordhaus, W. D. (1975). The political business cycle. *Review of Economic Studies, 42*, 169–190.
Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2019). Behavioral dishonesty in the public sector. *Journal of Public Administration Research & Theory, 29*, 572–590.
Ordeshook, P. (1986). *Game theory and political theory*. Cambridge Press.
Osborne, M. J., & Slivinski, A. (1996). A model of political competition with citizen candidates. *Quarterly Journal of Economics, 111*, 65–96.
Padovano, F. (2013). Are we witnessing a paradigm shift in the analysis of political competition? *Public Choice, 156*, 631–651.
Padovano, F., & Petrarca, I. (2014). Are the responsibility and yardstick competition hypotheses mutually consistent? *European Journal of Political Economy, 34*, 459–477.
Page, B. I. (1976). The theory of political ambiguity. *American Political Science Review, 70*, 742–752.
Paldam, M., & Nannestad, P. (2000). What do voters know about the economy?: A study of Danish data, 1990–1993. *Electoral Studies, 19*, 363–391.
Persson, T., & Tabellini, G. (2000). *Political economics*. MIT Press.
Pétry, F., & Collette, B. (2009). Measuring how political parties keep their promises: A positive perspective from political science. In: Imbeau L. (Ed.) *Do they walk like they talk?. Studies in Public Choice* (vol. 15). New York, NY: Springer.
Pettersson-Lidbom, P. (2008). Do parties matter for economic outcomes? A regression-discontinuity approach. *Journal of the European Economic Association, 6*, 1037–1056.
Pickering, A., & Rockey, J. (2011). Ideology and the growth of government. *Review of Economics and Statistics, 93*, 907–919.
Plane, D. L., & Gershtenson, J. (2004). Candidates' ideological locations, abstention, and turnout in U.S. midterm Senate elections. *Political Behavior, 26*, 69–93.
Potrafke, N. (2018). Government ideology and economic policy-making in the United States-a survey. *Public Choice, 174*, 145–207.
Praprotnik, K. (2017). Jurisdiction, time, and money: The role of resources in pledge fulfillment. *Party Politics, 23*, 848–859.
Reed, W. R., & Cho, J. (1998). A comparison of prospective and retrospective voting with heterogeneous politicians. *Public Choice, 96*, 93–116.
Roemer, J. (1994). A theory of policy differentiation in single issue electoral politics. *Social Choice and Welfare, 11*, 355–380.
Rogoff, K. (1987). *Equilibrium political budget cycles. Working Paper 2428*. National Bureau of Economic Research.
Rovny, J. (2012). Who emphasizes and who blurs? Party strategies in multidimensional competition. *European Union Politics, 13*, 269–292.
Sabatini, F. (2012). Who trusts Berlusconi? An econometric analysis of the role of television in the political arena. *Kyklos, 65*, 111–131.
Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics, 36*, 387–389.
Sanders, D., Clarke, H. D., Stewart, M. C., & Whiteley, P. (2011). Downs, stokes and the dynamics of electoral choice. *British Journal of Political Science, 41*, 287–314.
Santolini, R. (2008). A spatial cross-sectional analysis of political trends in Italian municipalities. *Papers in Regional Science, 87*, 431–451.
Schofield, N. (2003). Valence competition in the spatial stochastic model. *Journal of Theoretical Politics, 15*, 371–383.
Schultz, C. (1995). The politics of persuasion when voters are rational. *Scandinavian Journal of Economics, 97*, 357–368.

Sencer, S. D. (1991). Read my lips: Examining the legal implications of knowingly false campaign promises. *Michigan Law Review, 90*, 428–469.
Shepsle, K. A. (1972). The strategy of ambiguity: Uncertainty and electoral competition. *American Political Science Review, 66*, 555–568.
Shi, M., & Svensson, J. (2006). Political budget cycles: Do they differ across countries and why? *Journal of Public Economics, 90*, 1367–1389.
Stadelmann, D., Portmann, M., & Eichenberger, R. (2014). The law of large districts: How district magnitude affects the quality of political representation. *European Journal of Political Economy, 35*, 128–140.
Stokes, D. E. (1963). Spatial models of party competition. *American Political Science Review, 57*, 368–377.
Thomson, R. (2011). Citizens' evaluations of the fulfillment of election pledges: Evidence from Ireland. *Journal of Politics, 73*, 187–201.
Todorov, A., Mandisodza, A., Goren, A., & Hall, C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*, 1623–1626.
Tufte, E. (1978). *Political control of the economy*. Princeton University Press.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.
Usher, D. (2005). Assessing the citizen: Candidate model. *Public Choice, 125*, 43–65.
van Weelden, R. (2013). Candidates, credibility, and re-election incentives. *Review of Economic Studies, 80*, 1622–1651.
Weaver, R. K. (1986). The politics of blame avoidance. *Journal of Public Policy, 6*, 371–398.
Winer, S., & Ferris, J. (2022). *Political competition and the study of public economics (elements in public economics)*. Cambridge University Press.
Wittman, D. (1983). Candidate motivation: A synthesis of alternative theories. *American Political Science Review, 77*, 142–157.
Wittman, D. (1989). Why democracies produce efficient results. *Journal of Political Economy, 97*, 1395–1424.
Wittman, D. (1990). Spatial strategies when candidates have policy preferences. In J. Enelow, & M. Hinich (Eds.), *Advances in the Spatial Theory of Voting*. New York: Cambridge Press.
Wray, J. H. (1981). Rethinking responsible parties. *Western Political Quarterly, 34*, 510–527.
Zeckhauser, R. (1969). Majority rule with lotteries on alternatives. *Quarterly Journal of Economics, 83*, 696–703.

The Effect of Social Interaction and Cultural Consumption on Voting Turnout

Marco Ferdinando Martorana and Isidoro Mazza

1 Introduction

In the interesting paper "Wallenstein's Power Problem and Its Consequences", Manfred Holler & Barbara Klose-Ullman (2008) embark on the unusual attempt to convince theatregoers that learning from game theory could be beneficial. Once we heard Manfred Holler saying that reverting the usual logic could be illuminating. Following his advice, here we try to suggest that it could be worthwhile looking at the cultural capital of theatregoers to shed a light on an issue that game theory, assuming fully rational agents, has struggled to explain, namely voters' participation in elections.

Fully rational behavior implies that, with strictly positive voting costs, people would vote only when they are pivotal. In fact, from Downs (1957), each voter preferring a candidate, votes if and only if the expected utility from victory of the preferred candidate is higher than voting costs. But if individuals are rational, and voting is purely instrumental to obtain the preferred electoral outcome, voting turnout in large elections should be very low, because the probability of being pivotal approximates zero as the number of potential voters increases. However, this hypothesis is rejected by the frequent observation that voting is definitely more common than abstaining in democratic systems.

A substantial literature has provided several potential solutions to the voting paradox. Some approaches abandon the assumption of fully rational forward-looking voters and assume bounded rationality. Other models keep rationality but associate

M. F. Martorana · I. Mazza (✉)
Department of Economics and Business, University of Catania, Corso Italia 55, 95129 Catania, Italy
e-mail: imazza@unict.it

M. F. Martorana
e-mail: marco.martorana@unict.it

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_18

a benefit to the *act* of voting itself (expressive voting approach).[1] A different group of models, the group-based models, operate within the realm of full rationality and focus on the probability for a voter of being pivotal when he or she belongs to a group adopting a common behavior in voting.[2]

The basic logic behind group-based voting models is that the structure and dynamics emerging in a group, through internal norms and mobilization, may induce its members to conform to cooperative behavior, and, in the domain of politics, may eventually encourage voting participation. In fact, individuals are active in different types of groups, ranging from clubs to unions, political parties and religious groups. Such groups differ in their structure, as well as in the strength of members' identification with the group, of obligations, ties and social norms. One may ask whether group membership itself matters or if differences among types of groups imply different levels of members' involvement (and investment), in turn affecting cooperative behavior differently.

Such questions evoke the definition of *social capital* developed by Bourdieu (1986, p. 248), as "the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance and recognition—or in other words, to membership in a group". According to Bourdieu, the level of social capital of an individual is a function of the size of the network she can mobilize and on the level of economic and cultural capital of people she is connected to (p. 249). Such proposition conveys Bourdieu's idea of social capital as a collective-owned capital, resulting from a complex set of interactions of individuals within groups, in the form of trust, recognition, respect and gratitude, which result from a process of individual and collective investment. Although often criticized for its elusiveness (Prell, 2006), such definition of social capital emphasizes that group membership per se is not a sufficient condition for cooperation. In Bourdieu's view, a network of relationships is the product of individual and collective efforts to transform contingent relations in durable obligations. Already Jacobs (1961) indicated that the interactions among people leaving in the same neighborhood can trigger cooperative behavior and support the voluntary supply of public goods. Putnam (1993) links social capital to associational activities suggesting a positive role in economic performances, in contrast with Olson's (1982) theory that associations of individuals pursuing a specific interest are expected to have a detrimental impact on growth. Coleman (1988), on the other hand, describes social capital as built by relations of trust and consensual social norms. A substantial empirical literature has subsequently developed to verify the positive impact of (different types of) social capital and social norms on development, through the transmission of cooperative behavior and trust (Keefer & Knack, 2008).

[1] The main limit of this approach is its tautological evidence, as individuals end up voting when they feel they should vote. Other solutions, within the fully rational framework, predicting a positive level of turnout include the game-theoretical models (Krueger and Acevedo, 2008; Palfrey and Rosenthal, 1983; 1985), info-based models (Larcinese, 2007) and group-based models (Ulhaner, 1989; Feddersen, 2004; Feddersen and Sandroni, 2006; Fowler, 2005).

[2] Surveys of rational solutions are provided, among others, by Blais (2000), Mueller (2003) and Geys (2006).

Building on such insights, this paper aims at contributing to the existing literature on voting behavior by investigating the forms of social activities and interactions able to reinforce intra-group communication and a sense of obligation that may ultimately lead to some cooperative behavior, in this case voting participation, which is classically interpreted as a form of public good provision (Tullock, 1971). To do so, we consider the links between voting participation and various forms of group memberships. They require different intensities of obligations and individual 'investments' to stay in the group, which helps to reinforce social relationships. We would expect that social interaction per se would not be significant for collective action. In other words, the mere participation in a group that is low-demanding, with no specific ethical, political or cultural connotation (e.g. sport clubs) would not contribute to accumulate social capital in the same way as membership in a religious or political group. Moreover, latter groups have a hierarchical structure (leader and followers) that, according to models of group mobilization, may spur collective action in a group. Results confirm this hypothesis. Social interaction has no significant influence on voting participation in unlikely groups characterized by some deep bond, such as political or religious faith.

We also verify whether cultural expenditure is connected to voting participation. The reasons for this examination are twofold. Firstly, cultural consumption may favor social interaction and even group membership. Finocchiaro Castro et al. (2021) highlight the role of cultural sensitivity, which is positively related to cultural capital and influences the individual approval of the voluntary supply of a public good (i.e. private donations for culture), trigging a virtuous circle of cooperative behavior. Secondly, a recent strand of literature has studied and found evidence of a connection between cultural participation and political attitudes (McAndrew et al., 2020). Therefore, we expect that cultural consumption will be positively related with voting. Again, results confirm the hypothesis. Moreover, we find that *not any* kind of cultural expenditure is significantly related to voting. Theatre attendance has a significant impact while going to see a movie. We suggest that such a result can be further understood by referring to the concept of cultural capital in its "embodied state", as a process of individual accumulation that implies a continuous effort that costs time and can be viewed as a process of investment (Bourdieu, 1986). Theatre attendance, in this sense, represents in our opinion a better proxy for such a voluntary accumulation process than cinema attendance, since the latter can be included within the concept of mass consumption.

Finally, we consider the effects of a phenomenon, which may have disruptive effects on social capital, namely residential mobility. Moving to a different residence can weaken your bonds and participation in a specific group, at least in the short run. In the medium-long run old (or new) links to old (or new) groups of equal or different kinds are likely to form again. Results confirm this hypothesis showing that changing residence affects voting participation within the first year but not later.[3]

[3] Moving may imply registration delays. In the UK, voting offices make a yearly check about residence. This means that you may not be listed as a voter if you moved recently without informing the public office.

To test our hypotheses, we use data from the British Household Panel Survey (BHPS), a household-based longitudinal study of individuals living in Great Britain. Using the information on four national elections (1992–2005).

This study's contribution to the literature on the voting paradox is twofold. First, it is an attempt to improve our understanding of the social environment where group-based voting participation is more likely to emerge because of social capital. Second, it studies the types of cultural activities that are related to political participation, through cultural capital accumulation. In line with our expectations, results show that voting participation is positively and significantly related to social interactions in groups that are characterized by strong bonds, such as political or religious faith, and to some forms of cultural consumption such as theatre attendance, but not to mass cultural consumption such as cinema attendance. The paper is organized as follows. Section 2 discusses the relevance of group-based behavior. Section 3 shows the data used in the regression. Section 4 explains the estimated model. Section 5 presents the results of the empirical investigation. Section 6 concludes the paper with few comments about the analysis that has been conducted.

2 Group-Based Models of Voting, Social Capital, and Cultural Capital

One of the most interesting and promising attempts to solve the paradox is based on the analysis of individual behavior within formal as well as informal groups. Starting from Uhlaner (1989), group-based models represent a path explored to reconcile the theory with observed voting patterns. In Uhlaner (1989), groups are large enough to be pivotal and candidates do not share the same position in the political dimension. Morton (1991) develops the group-based approach by examining turnout equilibria in a strategic model with risk adverse voters. Voting as a strategic participation game incorporating groups is also analyzed by Schram and van Winden (1991) and Schram and Sonnemans (1996a, 1996b), which divide group members into opinion leaders and pressure consumers and consider both inter-groups and intra-groups correlations. As in former models, the basic intuition refers to the ability of groups to be pivotal in elections. Evidences of higher turnout rates associated with group membership and intra-group communication are provided in a laboratory experiment (1996b; Schram & Sonnemans, 1996a). Feddersen (2004) distinguishes between "group-based voting models of mobilization" and "group-based ethical voter models". Group-based ethical models assume instead that individuals are motivated to participate in elections by a sense of civic duty or ethical obligations (see also Feddersen & Sandroni, 2006) and by evaluations at aggregate level, as in traditional ethical models. Nevertheless, group membership effect might hide the influence of social interactions. For example, Fowler (2005) assumes that a single act of voting many individuals are linked together by social connections (turnout cascade effect) and shows that ideological homogeneity amplifies the turnout cascade effect. Mobilization models highlight the relations within a group, especially between leaders and

followers, to explain how leaders' efforts may determine high turnout levels among the group's members. Apart from the influence of a leader, mobilization depends on a group's internal cohesion, and on the strength of bonds within a group, which vary from group to group. In fact, the empirical literature has so far focused on some specific types of groups, characterized by a strong identity and/or a hierarchical internal structure, such as religious (Margolis, 2018), ethnic groups (Houle, 2018, 2019), and labor unions (Gray & Caul, 2000). Houle (2019) finds that within group homogeneity fosters members' identification with the group and positively affects ethnic voting.[4] Gray and Caul (2000) explain the fall in turnout rates in western countries with the decline of group mobilization of labor unions.[5]

Group-based collective action is also addressed by Bourdieu (1986) in his view of social capital as social obligations, connections, and networks available to an individual. However, the existence of a group is not sufficient per se to lead people to cooperation. The key aspect in Bourdieu's logic is that social capital generates from networks of relationships as the result of individual and collective investments transforming occasional relations into durable obligations. Such obligations may include cooperating in activities that affect group welfare, such as voting. Thus, the link between group membership and voting participation lies in individual investment, connected to the strength of group obligations, that is at the basis of social capital accumulation.

The same logic applies, in the Bourdieu's thought to cultural capital accumulation in its *embodied state,* which, again, implies a process of individual investment of a similar type. This is one of the reasons why we also study the effect of different types of cultural consumptions on voting participation. Also, cultural capital is connected to education, which has been acknowledged as a voting predictor (see Mueller, 2003). Finally, a recent strand of literature has found evidence of a connection between political values and attitudes, on the one hand, and cultural consumption (McAndrew et al., 2020) as well as working in the cultural sector (which is positively connected with cultural consumption as shown by Chan et al., 2020), on the other hand.

In a nutshell, we expect that not all the types of group membership, and not all the types of cultural activities affect voting participation. Regarding groups, we presume that only those groups that are characterized by strong internal ties will enforce social capital accumulation and, with respect to cultural consumption, we presume that only a subset of activities can be interpreted in terms of individual investment in cultural capital accumulation. Overall, results confirm our expectations. However, it is worth mentioning that we consider voting in general, without a specific indication of the policy platforms that are presented. If individuals were called to express their preferences about a specific policy, for example regarding culture (Holler & Mazza, 2013), the motivations to participate can differ according to the individual preferences regarding that kind of policy.

[4] Ethnic voting can be viewed as a group-based type of voting, which has been proven to be relevant in explaining voter behavior (Boudreau et al., 2019; Houle, 2018; Kalkan et al., 2018).

[5] See Kostelka and Blais (2021) for a discussion on turnout rates and group mobilization, and the related literature.

3 Data Description

Data for the analysis are from the BHPS. This is a longitudinal study of persons living in Great Britain based on household units. The panel consists of eighteen waves (1991–2009), including more than 10,000 individuals in its first wave.[6] The BHPS does not provide much information about political attitudes that are usually included in Political datasets. We use the subsample of those who are eligible[7] to vote in the electoral cycle 1997–2005 and we focus on UK General Elections.[8] BHPS includes only individuals who live in households while those who live in institutions are excluded and this can be considered the first possible source of bias.

According to Uhrig (2008), attrition occurs mainly between the first two waves while it is negligible in the rest of the panel set. However, as our research question refers to elections according to in-time characteristics there are no reasons for using information belonging to the first wave (i.e. 1991). In the following table we present variables description and summary statistics.

4 Model Specification

Consider the following generic probit model:

$$y_{it}* = \beta_1 y_{t-1} + \beta_2 x_i + \beta_3 D_{it} + \varepsilon_{it} \tag{1}$$

where $y_{it}*$ is individual latent pseudo-propensity to vote in general elections; y_{t-1} is the lagged dependent variable (i.e. the observed voter behavir at time $t-1$); x_{it} is a set of individual characteristics; D_{it} is a set of dummies indicating groups' membership and leisure activities; and ε_{it} is the random component. An individual votes if her pseudo-propensity is positive: $y_i = 1$ if $y_i* > 0$. We estimate the models as pooled probit and allow for observations to be correlated within households.

We estimate a set of 4 models to study the connection between group membership and activity, and voting. In model 1, we estimate a null model that does not include any D variable. In model 2, we add three variables indicating group membership. We consider religious groups and trade unions as hierarchical groups, and sports clubs as informal groups. In model 3, we include a set of dummies indicating individual leisure activities. We consider the attendance to theatrical representation, voluntary unpaid activities, the attendance to cinema, and a dummy indicating whether respondents frequently get out with friends for dinner. Finally, in model 4, we aim at considering the effect of residential mobility on turnout. Purposely we add two other variables:

[6] The number of surveyed households and, thus, the number of individuals increased through waves, as additional national samples (Northern Ireland, Wales and Scotland) were added.

[7] We use self-reported eligibility to vote, and anyway drop observations of individuals with age lower than 18 (the eligibility age to vote).

[8] We still use information on the 1992 election to compute the lag of electoral participation.

moved is a dummy taking value 1 if respondent moved to the present address in the last year. The second variable *l-moved* indicates if respondent had moved to the present address at t-1.

As controls, we use educational level dummies, job status (6 dummies), region (3 dummies), marital status (6 dummies), gender, class of age and self-reported interest in politics (see Table 1 for variable description).

We perform the usual link test for specification and the Hosmer-Lemenshow (2000) goodness-of-fit test. We also perform a Box-Tidwell (1962) estimation to check if any predictor transformations are needed. A test on the random effects estimation confirms the assumption of no correlation across observations for each individual.[9] Finally, we re-estimate the models by using survey weights in order to check if the results hold (Table 2).

5 Results

We discuss the outcomes in Table 2 in terms of group membership effect and leisure activity effect. Model 1 is the baseline model that shows a positive effect of education and age on electoral participation, as commonly found in the literature. The positive and significant coefficient of education also supports the idea that cultural capital accumulation increases individual propensity to cooperative behaviour as in Finocchiaro Castro et al. (2021). In model 2, we add group membership in the model specification, finding that only *religious group* and *union membership* are statistically significant. On the contrary, being a member of a sports club does not affect turnout propensity. We interpret such a result by considering the higher intensity of obligation related to hierarchical groups than informal groups. As we expected, low-demanding groups do not affect cooperative attitudes or political participation. Religious groups and unions are characterized by political and cultural connotations as well as by a hierarchical structure strengthening social interactions and favoring mobilization.

The same logic applies to model 3, where we also consider leisure activities. In this case, we would expect all the leisure variables to have an effect on voting propensity if the hypothesis of social interactions holds. Here, only *attendance to theatre* and *voluntary* (unpaid) *activities* have a positive and statistically significant effect on turnout probability. Our interpretation of these results is twofold: on the one side, we suggest that the positive and significant effect of voluntary activities highlights the role of prosocial behavior that is coherent with cooperative attitudes. On the other side, the positive effect of attendance to theatre underlines the existence of social capital and cultural capital as a result of a process of accumulation and embodiment that cannot be reduced just to the level of education.

[9] Models 2, 3 and 4 pass specification and goodness of fit test. Test outcomes, Box-Tidwell estimations, correlation matrix, weighted and subsample estimations can be provided upon request.

Table 1 Summary statistics and variable descriptions

Variable	Obs	Mean	Std. dev	Min	Max	Description
Turnout	19,085	0.752371	0.431646	0	1	Dummy variable taking value 1 if respondent voted at last general election and 0 otherwise
Lag turnout	19,085	0.775164	0.417485	0	1	The lag of variable turnout
Sport club	19,085	0.181975	0.385834	0	1	Dummy variable taking value 1 if respondent is a member of a sport club
Religious group	19,085	0.133089	0.33968	0	1	Dummy variable taking value 1 if respondent is a member of religious group
Union membership	19,085	0.185958	0.389083	0	1	Dummy variable taking value 1 if respondent is a member of a trade or labor union
Moved	19,085	0.083416	0.276518	0	1	Whether a respondent moved from a place to another in the last year
Theatre	19,085	0.358868	0.479681	0	1	Dummy taking value 1 if respondent declared to attend theatrical representation several times a year
Cinema	19,085	0.456275	0.498098	0	1	Dummy taking value 1 if respondent declared to attend cinema several times a year
Voluntary activity	19,085	0.161331	0.367846	0	1	Dummy taking value 1 if respondent declared to attend voluntary not paid activities several times a year
Eat out	19,085	0.88656	0.317138	0	1	Dummy taking value 1 if respondent declared to eat out several times a year
Education	19,085	2.039455	0.581328	1	3	A set of three dummies: *high education* (ISCED 5–6), *intermediate* (ISCED 3–4), *low education* (ISCED 0–2)
Marital status	19,085	2.069112	1.767446	1	6	Set of dummies indicating marital status

(continued)

Table 1 (continued)

Variable	Obs	Mean	Std. dev	Min	Max	Description
Female	19,085	1.548232	0.497681	0	1	Gender
Job status	19,085	2.58601	1.328509	1	5	Set of dummy variables indicating respondent job status: self-employed, in paid employ, unemployed, retired, full-time student, other
Region	19,085	2.014986	0.82305	1	3	Set of dummies
Ethnic group	19,085	1.05727	0.366428	1	4	Set of dummies
Age	19,085	45.44228	12.81782	22	70	A set of three class of age dummies (age < 30; 30 < age < 65; age > 65)
Interest in politics	19,085	2.315274	0.891164	1	4	Set of dummies indicating respondent's self-reported level of interest in politics

Source Our computation

The attendance at the theatre shows the individual investment in such a process. One may argue that going to watch a movie has a monetary cost as well as going to theatre. Nevertheless, the former represents a type of mass consumption that cannot be included in such a work of self-acquisition and personal investment presupposing costs that Bourdieu (1986) associates with cultural capital accumulation.[10]

Finally, in model 4, we study the effect of residential mobility on turnout probability. Results confirm our expectations: probability to vote is lower if the voter moves to a new house during an election year. This effect disappears after one year. We would interpret this result by arguing that residential mobility weakens social interactions and so affects cooperative behavior. However, the cost of voting now includes the cost (also in terms of time) of registration on the electoral rolls. In United Kingdom, local electoral offices deliver registration forms each year (between May and November) to every house to maintain the electoral registry. Election takes place generally in May, hence it is technically possible that if a voter moves during an election year she is not registered on the roll.

[10] Differences in theatre and cinema attendance could be partially related to income level, as theatres are generally much expensive and also less accessible. In fact, their density is lower than the one of cinemas, especially in the countryside (meaning that people have to move to closest town, increasing the true cost of attending a piece). The BHPS does not include direct measures of income but some self-reported measures of economic condition, which would work as proxies for income level if not largely unreliable and inconsistent. We also choose not to include them as controls to avoid inference problem due to correlation, given that we are interested in controlling for educational level. However, we believe that using job status dummies jointly with education level dummies allows us to effectively control for differences in cultural participation due to income level.

Table 2 Estimation outcomes

	(1)	(2)	(3)	(4)
Variables	Turnout	Turnout	Turnout	Turnout
Sport club		0.007	−0.003	−0.004
		(0.030)	(0.031)	(0.031)
Religious group		0.174***	0.137***	0.137***
		(0.037)	(0.038)	(0.038)
Union membership		0.179***	0.177***	0.169***
		(0.032)	(0.032)	(0.0322)
Moved				−0.442***
				(0.042)
Moved at t-1				0.004
				(0.038)
Eat out			−0.037	−0.036
			(0.037)	(0.037)
Theatre			0.096***	0.101***
			(0.026)	(0.026)
Cinema			0.005	0.013
			(0.026)	(0.026)
Voluntary activity			0.116***	0.118***
			(0.034)	(0.034)
High educated	0.156***	0.128***	0.105***	0.129***
	(0.036)	(0.036)	(0.037)	(0.038)
Low educated	−0.061**	−0.038	−0.019	−0.015
	(0.029)	(0.029)	(0.030)	(0.030)
Age < 30	−0.259***	−0.246***	−0.234***	−0.197***
	(0.028)	(0.028)	(0.028)	(0.029)
Age > 65	0.233***	0.231***	0.232***	0.223***
	(0.051)	(0.051)	(0.051)	(0.051)
Lag turnout	1.144***	1.135***	1.133***	1.135***
	(0.026)	(0.026)	(0.026)	(0.026)
Constant	0.177	0.120	0.101	0.117
	(0.150)	(0.151)	(0.153)	(0.156)
Other controls (see Table 1)	Yes	Yes	Yes	Yes
Observations	19,085	19,085	19,085	19,085
Pseudo r^2	0.252	0.254	0.255	0.261

(continued)

Table 2 (continued)

	(1)	(2)	(3)	(4)
Variables	Turnout	Turnout	Turnout	Turnout
chi2	4381	4408	4411	4436
Link test	Yes	Yes	Yes	Yes
Hosmer Lemenshow	Yes	Yes	Yes	Yes

Source Our computation. *Note* Clustered (by households); standard errors in parentheses; ***p < 0.01, **p < 0.05, *p < 0.1

6 Concluding Comments

Our study should be interpreted as an analysis of the relationship between social and cultural capital and a specific collective action, namely participation to vote. We have shown that people connected through networks that demand a high level of individual engagement and impose social obligations are more likely to vote. Moreover, the hypothesis that accumulation of cultural capital through adequate investments in education is expected to make people more willing to interact socially and more sensitive to cooperative behavior (Finocchiaro Castro et al., 2021) is confirmed by the significantly positive relationship between cultural capital and voting participation. Interestingly, the empirical analysis indicates that the *levels* of investment are crucial to find a significant effect of social and cultural capital on voting. Activities requiring a low engagement do not seem to play a role. The political implications of this study's findings are of some relevance. Common wisdom says that the 'health' of a democracy firstly depends on voting participation. People's detachment from politics reduces political accountability and fosters populism and extremism, as we are increasingly observing in western countries. The message of this paper is that social and cultural capital have a crucial role in preserving democracy powered by voting.

Results however encourage further analysis, in our view. In particular, we find clear evidence that social capital built through investments in group activities has a significant role in motivating participation especially when these groups have an established hierarchical structure. This result supports the hypothesis derived by group mobilization models and it is also consistent with the definition of social capital by Bourdieu (1986). Also, his definition of cultural capital offers an insightful interpretation of the observed difference in voting participation between theatre and movie attendances. Further efforts should be devoted to better define the behavioral dynamics within the group to reach more precise and testable hypotheses.

References

Blais, A. (2000). *To vote or not to vote?: The merits and limits of rational choice theory*. University of Pittsburgh Press.

Bourdieu, P. (1986). The forms of social capital. In J. E. Richardson (Ed.), *Handbook of theory of research for the sociology of education* (pp. 241–258). Greenwood Press.

Boudreau, C., Elmendorf, C. S., & MacKenzie, S. A. (2019). Racial or spatial voting? The effects of candidate ethnicity and ethnic group endorsements in local elections. *American Journal of Political Science, 63*(1), 5–20.

Box, G. E., & Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics, 4*(4), 531–550.

Chan, T. W., Henderson, M., Sironi, M., & Kawalerowicz, J. (2020). Understanding the social and cultural bases of Brexit. *The British Journal of Sociology, 71*(5), 830–851.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*(Supplement), S95–S120.

Downs, A. (1957). *An economic theory of democracy*. Harper.

Feddersen, T. J. (2004). Rational choice theory and the paradox of not voting. *The Journal of Economic Perspectives, 18*(1), 99–112.

Feddersen, T., & Sandroni, A. (2006). A theory of participation in elections. *The American Economic Review, 96*(4), 1271–1282.

Finocchiaro Castro, M., Mazza, I., & Romeo, D. (2021). The role of cultural capital on the voluntary contributions to cultural goods: A differential game approach. *Games, 12*(1), 27.

Fowler, J. H. (2005). Turnout in a small world. In A. Zuckerman (Ed.), *Social logic of politics* (pp. 335–344). Temple University Press.

Geys, B. (2006). 'Rational' theories of voter turnout: A review. *Political Studies Review, 4*(1), 16–35.

Gray, M., & Caul, M. (2000). Declining voter turnout in advanced industrial democracies, 1950 to 1997: The effects of declining group mobilization. *Comparative Political Studies, 33*(9), 1091–1122. https://doi.org/10.1177/0010414000033009001.

Holler, M. J., & Klose-Ullmann, B. (2008). Wallenstein's power problem and its consequences. *Czech Economic Review, 2*(3), 197–218.

Holler, M. J., & Mazza, I. (2013). *Cultural heritage: Public decision-making and implementation*. Edward Elgar Publishing.

Hosmer, D. W., & Lemenshow, S. (2000). *Applied logistic regression*. Wiley.

Houle, C. (2018). Does ethnic voting harm democracy? *Democratization, 25*(5), 824–842.

Houle, C. (2019). Religion, language, race and ethnic voting. *Electoral Studies, 61*, 102052.

Jacobs, J. (1961). *The death and life of great American cities*. Vintage Books.

Kalkan, K. O., Layman, G. C., & Green, J. C. (2018). Will Americans Vote for muslims? Cultural outgroup antipathy, candidate religion, and US voting behavior. *Politics and Religion, 11*(4), 798–829.

Keefer, P., & Knack, S. (2008). Social capital, social norms and the new institutional economics. In *Handbook of new institutional economics* (pp. 701–725). Berlin: Springer.

Kostelka, F., & Blais, A. (2021). The generational and institutional sources of the global decline in voter turnout. *World Politics, 73*(4), 629–667.

Krueger, J. I., & Acevedo, M. (2008). A game-theoretic view of voting. *Journal of Social Issues, 64*(3), 467–485.

Larcinese, V. (2007). The instrumental voter goes to the newsagent: Demand for information, marginality and the media. *Journal of Theoretical Politics, 19*(3), 249–276.

Margolis, M. F. (2018). How far does social group influence reach? Identities, elites, and immigration attitudes. *The Journal of Politics, 80*(3), 772–785.

McAndrew, S., O'Brien, D., & Taylor, M. (2020). The values of culture? Social closure in the political identities, policy preferences, and social attitudes of cultural and creative workers. *The Sociological Review, 68*(1), 33–54.

Morton, R. B. (1991). Groups in rational turnout models. *American Journal of Political Science*, 758–776.
Mueller, D. C. (2003). *Public choice III*. Cambridge University Press.
Olson, M. (1982). *Rise and decline of nations: Economic growth.* Yale University Press.
Palfrey, T. R., & Rosenthal, H. (1983). A strategic calculus of voting. *Public Choice, 41*(1), 7–53.
Palfrey, T. R., & Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review, 79*(1), 62–78.
Prell, C. (2006). Social capital as network capital: Looking at the role of social networks among not-for-profits. *Sociological Research Online, 11*(4), 39–52.
Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton.
Schram, A., & Sonnemans, J. (1996a). Voter turnout as a participation game: An experimental investigation. *International Journal of Game Theory*, *25*(3), 385–406.
Schram, A., & Sonnemans, J. (1996b). Why people vote: Experimental evidence. *Journal of Economic Psychology*, *17*(4), 417–442.
Schram, A., & Van Winden, F. (1991). Why people vote: Free riding and the production and consumption of social pressure. *Journal of Economic Psychology, 12*(4), 575–620.
Tullock, G. (1971). Public decisions as public goods. *Journal of Political Economy, 79*(4), 913–918.
Uhlaner, C. J. (1989). Rational turnout: The neglected role of groups. *American Journal of Political Science*, 390–422.
Uhrig, N. S. C. (2008). The nature and the causes of attrition in the British household panel survey. *ISER working Paper*, Colchester: ISER.

The Political Economy of Buchanan's Samaritan's Dilemma

Alain Marciano

1 Introduction

The samaritan's dilemma is a situation that has largely been studied in economics and social sciences. It has been shown that it occurs in a large number of situations (Boettke & Martin, 2010; Boone, 1996; Bovard, 1983; Burnside & Dollar, 2000; Coyne, 2008, 2013; Futagami et al., 2004; Rajan & Subramanian, 2005; Skarbek, 2016; Stone, 2008; Wilson, Andersson, Ostrom and Shivakumar, 2005; Wagner, 2005; Williamson, 2010). It is, therefore, an important phenomenon. However, the reference to the economist who—as it seems—used the term for the first time, James Buchanan, is frequently lost. This paper focuses on the role the Samaritan's dilemma played in Buchanan's work.

It was in an essay published in 1975 in a collective volume edited by Edmund Phelps, entitled *Altruism, Morality, and Economic Theory* (1975) that Buchanan put forward and analyzed the problem Samaritans could face when they help people in need by, for instance, transferring them money. The dilemma is a trap in which Samaritans are caught when and because the persons they help do not react to the help received and make no effort to reciprocate it. Now, Buchanan believed or started from the assumption that a Samaritan expects that her help, gift or transfer will lead

A. Marciano (✉)
University of Montpellier and MRE, Montpellier, France
e-mail: alain.marciano@umontpellier.fr

Faculté d'Économie, Université de Montpellier, Avenue Raymond Dugrand, 79606, F-34960 Montpellier Cedex 2, CS, France

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_19

the recipient to change his behavior, make some additional efforts, in particular, "work".[1] Therefore, the Samaritan cannot but be dissatisfied with the lack of effect the transfer has on the recipients' behavior. To avoid this asymmetrical, unequal—and unpleasant—situation in which she helps and the recipient does not make any effort in return, the Samaritan could decide to stop behaving charitably and stop helping the potential recipient. Choosing this course of action would suppress the Samaritan's dissatisfaction with seeing her charity not reciprocated and also, in Buchanan's mind, would incite the recipient to change his behavior and work. Symmetry and equality would thus be restored. But, according to Buchanan, not helping people in need is costly. The Samaritan would suffer from seeing the recipient starve and from the loss of utility that stems from the impossibility of behaving benevolently. An unpleasant situation replaces another one. Hence, the dilemma the Samaritan faces: either she helps someone who does not return her help or she does not help but suffers from the cost of not helping.

To Buchanan, Samaritans could prevent being caught in the dilemma if and only if they were able to follow an ethic of individual responsibility and exhibit what he called "strategic courage"—that is the courage *not* to help others even despite the costs it implies for the Samaritan. That was the responsibility Samaritans had and should face. But Buchanan doubted they would. The "cliche that modern man has 'gone soft'" (Buchanan, 1975, 75) was only too true. Samaritans could not resist their benevolence and the demands of recipients whom Buchanan "labelled… parasite[s]" (1975, 75, e.g.). They were "incapable of making the choices that are required to prevent [their] exploitation by predators of [their] own species." (74) Buchanan insisted: "What we may call 'strategic courage' may be a markedly inferior economic good, and what we may call 'pragmatic compassion' may be markedly superior" (1975, 75). To Buchanan, individuals were afraid to behave as responsible Samaritans, to assume their responsibilities.

These particularly negative statements can be explained by what was happening in the U.S.A. in the late 1960s and early 1970s when Buchanan started to draft his essay. He was extremely pessimistic about the situation in his country and in Western countries in general. It was this pessimism that had led him to write "The Samaritan's dilemma" (see). Yet, a product of its times, "The Samaritan's dilemma" also perfectly fits Buchanan's view on public economics and public choice or, put differently, his views on political economy. This is what we argue in this essay. More precisely, our goal is to show why situations like the ones described in the 1975 essay are perfectly consistent, and even the consequence of Buchanan's analysis on cooperation in the provision of public goods and the removal of externalities. In other words, Samaritan's dilemmas are inherent to how Buchanan envisaged cooperation

[1] In his description of the interaction between the Samaritan and the recipient, Buchanan wrote that the recipient had two strategies: "the potential parasite… may work … Or… refuse work" (1975, 76). Then, the many examples he gave of Samaritan dilemmas evidence that it was not only a matter of "work". For instance, he used the case of a mother who hesitates to "spank a misbehaving child". Indeed, it hurts her to have to punish her child but "spanking may be necessary to instill in the child the fear of punishment that will inhibit future misbehavior." (76) Here, it is no longer a matter of "work" strictly speaking. it seems to be more of a matter of changing the behavior of the child.

and defection around public goods and externalities. More globally, the Samaritan's dilemma is part of Buchanan's objective of developing a form of non-Samuelsonian welfare economics.

Our demonstration rests on Buchanan's claim that individuals cooperate in small groups—they spontaneously internalize the external consequences of their actions—but do not cooperate in large groups. What we show, and this is original, is that Samaritan's dilemmas arise in what can be called "intermediary" situations, that is, when individuals behave as if they were in a small group while they are in a large one, or the reverse. This is the first element we put forward. Then, this is the second aspect of our analysis, Buchanan was perfectly aware that this kind of situation could occur. He acknowledged it at the end of the 1950s but he did not take the situation seriously before the end of the 1960s. This is also what we show in this paper.

The originality of our analysis lies in the fact that it has never been shown that the Samaritan's dilemma is consistent with Buchanan's analysis of cooperation. This finding is interesting to shed new light on Buchanan's work. That is not all. Our analysis is useful also because of the role of "numbers" in the occurrence of the dilemma. Indeed, if it is true that our analysis means that the scope for cooperation is narrower than what Buchanan himself acknowledged, it also reinforces the importance of relying on small groups to guarantee cooperation among individuals.

2 Market Failures, Cooperation, and Numbers

One of the most frequently admitted claims in economics is that, as soon as interdependencies exist between individuals, markets fail to allocate resources efficiently. This is, therefore, the case when there are public goods or spillover—external—effects. Indeed, individuals being rational and self-interested have no incentive to reveal their preferences—their willingness to pay—for the public good or to internalize the costs of their actions on others. This was, in particular, the view held by Richard Musgrave (1939, 1959)—who spoke about "the absence of a general willingness to comply with the obligation to contribute" (1939, 220). This was also how Paul Samuelson concluded his article on "The Pure Theory of Public Expenditure": "It is in the selfish interest of each person to give false signals, to pretend to have less interest in a given collective consumption activity than he really has" (1954, 388–389).

Buchanan for hist part wanted to develop a non-Samuelsonian perspective in economics. He himself suggested that it was his goal in a comment he sent to Samuelson in 1955 about the famous paper Samuelson had published in 1954 (cf. Boettke et Marciano, 2020; Marciano, 2021). Buchanan meant that he refused the use of a social welfare function to determine the price individuals should pay for public goods (or external effects). To him, no external observer could know what were the individual preferences for public goods, and therefore, could say what was their marginal willingness to pay. Consequently, prices such as those based on the

use of a social welfare function would in all likelihood be different from those individuals would be willing to pay. To avoid that coercive move, Buchanan suggested relying on what individuals want to pay for the public goods they consume, or for the external effects their behaviors generate.

The suggestion made sense only because Buchanan—and, beyond, also the economists from the Virginia School of Political Economy—was convinced that individuals reveal their true preferences, and therefore, accept to contribute to the provision of the public goods they consume or internalize the effects their actions have on others *even if they are rational and motivated by their self-interest* (Marciano, 2016). A conviction Buchanan held for a long time (Marciano, 2013). And, when he started to admit that individuals could not contribute to the provision of public goods, Buchanan only partially abandoned his confidence in individuals. Voluntary contribution remained the rule. A condition had nonetheless to be satisfied to guarantee that individuals would cooperate: that the group of individuals involved in the interaction be rather small. Thus, one may note, Buchanan put explicitly forward his first arguments on the role of the size of groups in the early sixties, that is before Mancur Olson (1965).

Thus, as the size of the group increases, individuals more and more follow their self-interest in the narrowest sense of the word, that is: they do not take into account the welfare of others. As Buchanan noted in a rather early article, "the individual's interest in the welfare of his fellow citizen falls off sharply as the group is enlarged." (1961a, 340). Therefore, "as the size of the group increases, any tacit adherence to moral or ethical principles that might inhibit individual utility-maximizing behaviour becomes more difficult to secure" (ibid.; see also, among others, Buchanan & Tullock, 1962; Buchanan & Kafoglis, 1963). Individuals cooperate less as the size of the group increases. This is what Buchanan called a "probabilistic" theory of free riding (1965b, 1968a, 85–88). How do cooperation and free riding relate to the size of groups?

3 Different Strategies in Different Types of Groups

To Buchanan, individuals' choice to cooperate or not to cooperate unsurprisingly depends on the gains and costs they expect to receive and incur from their behavior. Those gains then obviously depend on how others behave. More precisely, since no one cannot know in advance how others will behave they depend on each individual's "own predictions about the behavior of others" (1965b, 3). Indeed, it is a matter of probability—"[t]he expected values depend, of course, on the probabilities that the individual assigns to the various patterns of behavior for 'others' than himself" (5) or, Buchanan repeated almost word for word, depends on the "probabilities assigned [by the individual] to each of the possible behavior patterns of others" (1968a, 85). These probabilities in turn depend on the impact or influence each individual believes his or her behavior will have on others. Individuals thus cooperate after having contemplated whether or not their behavior will "exert some influence on the behavior of others in the group" (1968a, 86). However, the effect such a belief will have on

cooperation can be positive or negative. An individual "may behave cooperatively hoping that his… "rivals" will emulate his… action" (Buchanan, 1967, 121). Or, by contrast, he may decide not to cooperate because he expects that "his own action in contribution will lower, not increase, the probability of others' making contributions of their own" (Buchanan, 1968a, 86). To have a definite answer, we need to take into account the size of the group.

Buchanan distinguished between two types of environments—large and small number environments. In large groups, individuals do not anticipate that their own behavior will influence the behavior of others—"only when the size of the group is critically large. Only in such large groups will the individual consider his own action to exert substantially no effect on the actions of others" (1965b, 9). As a consequence, individuals behave non-strategically, "simply react[ing] or adjust[ing] to the behavior of "others" in a manner similar to his reaction to natural environment" (1967, 113). "Natural" refers to the fact that individuals treat others as part of *nature*, taking their behavior as given and assuming that how others behave is independent of their own behavior—"The behavior of the other is embodied as data in the choice calculus, but the other person is not considered to be subject to influence or control, positively or negatively" (Buchanan, 1967, 111). Therefore, individuals behave as if they were independent of others. They ignore the interdependencies that link them to others. They simply maximize their own, private, utility. Or, as Buchanan put it, they follow "the private maxim" or "the expediency criterion" (1965b, 2).

Now, since the decision to maximize one's utility, that is to follow the expediency criterion or the private maxim that is, comes from the belief that each individual is independent from others, one understands that each individual makes this choice by ignoring what others do. Whether or not others cooperate, each individual privately prefers not to cooperate—"[i]n a group of critically large size, the individual will tend to adopt the rule of following the expediency criterion even if he thinks that *all* of his fellow citizens are saints" (1965b, 7; italics in original). Even if they understand that unconditional defection is not the best strategy and that would be better off by cooperating with others. Convinced that they are not able to influence the behavior of others, no one will change his own behavior. No reason can lead individuals to choose to behave differently—"Rationally, he cannot adopt the moral law as a principle for his own behavior" (1965b, 7). Or, to put it in terms that will be useful for the rest of the analysis, individuals have a dominant strategy. They are unconditional defectors. Buchanan was rather clear about that: "The individual in a large-group, public-goods interaction… face[s]… no pressure or incentive to behave cooperatively", because he behaves independently from others (1967, 121). As a consequence, the individuals who are in large groups are trapped in what Buchanan called the "large number dilemma" (1965b), which is a form of prisoner's dilemma: "The dilemma is a real one, and it is similar to, although not identical with, that which is commonly discussed in game theory as "the prisoners' dilemma" (Buchanan, 1965b, 8). This situation corresponds to what Samuelson, and other economists, viewed as the standard case of market failure.

Let us note here that, to Buchanan, "The individual is caught in a dilemma by the nature of his situation; he has no sensation of securing benefits at the expense of others in any personal manner" (1968a, 83). Not cooperating, following the expediency criterion, means that the individuals maximize their own utility without taking into account or acknowledging the presence of others. Each of them behaves as if others were not there. In other words, choosing not to cooperate *in a large group* was not interpreted by Buchanan as meaning that individuals would free ride. From this perspective, Buchanan found the terminology used in public economics about free riding "misleading" (1968a, 83). Indeed, free riding "suggests some deliberate effort on the part of the choosing individual to secure benefits at the expense of his fellows" (Buchanan, 1965b, 9). It implies that individuals try "to shift a major share of the burden onto the other while himself securing a share of the benefits" (1967, 114). They really adopt anti-social behaviors (see Marciano, 2015). Or, to use another of Buchanan's words, they try to "exploit" others. In other words, a free rider acknowledges and takes into account others, which is incompatible with how individuals are supposed to behave in large groups. Free riding is more a behavior that could occur, but doesn't, in small groups.

In small groups such as an "isolated setting" (Buchanan, 1965b, 6) of 3 persons or "a desert island" (1965b, 6) or when "personal interaction is recognized" (Buchanan, 1968a, 86), the situation is different. Indeed, "utility maximization in a small number setting will not exhibit the observable properties of utility maximization in a large number setting" (Buchanan, 1978, 366). More precisely, individuals no longer follow their "narrowly defined self-interest" (Buchanan, 1978, 366) but rather adopt "moral or ethical principles" (Buchanan, 1961a, 340). Individuals adapt their behavior to the behavior of others or, to use Buchanan's words, they behave "strategically" (1968a, 91). They no longer ignore the interdependencies that link them to others. Indeed, by contrast with what happens in large number environments, individuals are aware that their behavior matters and may influence others: each individual "will tend to recognize that his own choice of a rule, and subsequent adherence to it, will to some considerable extent influence the similar choices to be made and followed by the other two members" (1965b, 6. Or, "So long as the interaction is limited to small groups, [the individual] will recognize that his own action can exert some influence on the behavior of others in the group" (1967, 115).

In particular, an individual who "contributes nothing… may assess the probability of noncooperation on the part of others higher than if he contributes some share. This change alone may be sufficient, on rational grounds, to cause him to contribute" (Buchanan, 1968a, 86). In other words, individuals cooperate because they know and anticipate that this behavior will lead others to cooperate too (and that defection would probably lead them to defect too). It thus seems reasonable to claim that, to Buchanan, the individuals who are in a small group or who are not narrow utility maximizers are unconditional cooperators, as long as they are in a small group or among other ethical individuals. They have a dominant strategy that consists of cooperating with the provision of public goods or internalizing the externalities their behavior can generate. Therefore, the large number dilemma disappears. Individuals

cooperate, contributing to the provision of public goods or internalizing the external effects of their actions on others. Even free riding does not exist.

4 When is There a Samaritan's Dilemma?

Thus, in Buchanan's analysis of free riding and cooperation, two main possibilities are put forward: the generalized prisoners' dilemma of large groups and the cooperative situation of small groups. What about intermediary situations? What if in a group, whether large or small, some individuals follow the expediency criterion and others adopt the ethical rule of conduct?

Before examining how Buchanan came to acknowledge this possibility, let us see where the analysis discussed in the previous section leads us. The outcome of an interaction between individuals who do not follow the same rule of behavior is rather complex. One would have understood that individual behaviors depend on whether or not they behave independently from others or strategically. And, in addition, adopting one behavior or the other depends on the size of the group. Thus, the outcome of an asymmetric interaction depends on whether a narrow maximizer finds himself in a small group or if an ethical behavior finds herself in a large group.

A narrow maximizer by definition ignores the interdependencies with others. As indicated above, this kind of individual does not change his behavior, even if others cooperate, because he does not expect that not cooperating (or cooperating, for that matter) could change the behavior of others. When he enters a small group and interacts with cooperators, and *when* he recognizes that there are cooperators around him, such an individual could be expected to cooperate. The fear of being sanctioned and excluded from the group should lead him to change his behavior. Then, the outcome of the game would be a cooperative situation, where both players cooperate. But there are cases in which the narrow maximizer could decide not to change his behavior: if or as long as he ignores others or believes that he cannot influence them or, and this is very important, if he realizes that the other members of the group might not be able to sanction and exclude him. After all, the costs of exclusion could be so high that narrow maximizers could not be excluded and remain in small groups. In that case, the narrow maximizer defects. It can even be said that he free rides and exploits the cooperator. Therefore, the outcome is the same as the asymmetric one described in Buchanan's "The Samaritan's dilemma" where the Samaritan of Buchanan's article is the individual who follows the small number environment rule of behavior and the recipient is the narrow maximizer.

The ethical individual entering a group of narrow maximizers faces a similar situation. He cooperates as long as he does not realize that the individual with whom he interacts does not behave strategically. The other, being a member of a group of narrow maximizers, defects. Therefore, the outcome corresponds to a Samaritan's dilemma. However, the dimension of free riding and exploitation does not exist. The narrow maximizers do not behave strategically. They do not try to shift their share of burden of the public good onto the ethical individual. The latter is nonetheless the

only one who cooperates in a world of defectors. This may not last. When the ethical individual finally understands the new situation, his behavior changes and he switches to the noncooperative strategy. The interaction between the two individuals becomes that of the prisoner's dilemma that already prevailed in the group. But again, the outcome of the interaction depends on the rapidity with which the individual adjusts to the new environment and on the costs of adjustment too. One would remind that, as said in the introduction of this paper, an important assumption made by Buchanan in the 1975 paper was that Samaritans were incapable of changing their behavior and of *not* behaving charitably because of their "softness".

Therefore, a Samaritan's dilemma exists as long as neither the ethical individual nor the narrow maximizer realizes that the nature of the interaction has changed and that they no longer interact with someone who adopts the same behavior as they do. The outcome of the interaction remains that of a Samaritan's dilemma as long as the "confusion" lasts or as long as individuals find it not advantageous enough or too costly to change their behavior. In other words, Samaritan dilemmas are one of the three possible outcomes that can result from individual interactions in Buchanan's theory of cooperation based on numbers and the size of groups or, put differently, from his probabilistic theory of free riding.

5 Toward the Samaritan's Dilemma

That such dilemmas could arise was perfectly clear in Buchanan's work before he started to write explicitly on cooperation and the size of groups and, consequently, before he wrote on the Samaritan's dilemma. Yet, it took him some time before admitting that such a situation was possible and durable.

What seems to be the first instance of a description of a Samaritan's dilemma can be found in the comment Buchanan made on the papers presented by Musgrave and, more interestingly, by Charles Tiebout at a 1959 conference in public finance he organized (see Marciano, 2013; Boettke, Marciano and Stein, 2021). Tiebout had discussed a situation that corresponds to the one described above. Newcomers arrive in a group in which a public good is already provided. They thus benefit from the good without having to contribute to its provision. Tiebout then put forward an institutional solution to deal with the problem. To him, zoning laws could be used to exclude these potential users from the consumption of the public good they did not have paid for (1961, 94; see also Tiebout, 1956, 420).

Buchanan did not disagree with the principle. Such laws or restrictions aiming at "[p]rohibition on entry" (1961b, 129) could indeed allow early settlers "to create a structure of property rights in 'taxpayers' surplus'" (129). In other words, the individuals who were financing the provision of public goods could exclude newcomers—who were supposed not to contribute to the provision of the goods in question—to guarantee their surplus. But, he added, exclusion was costly. Indeed, to exclude the newcomers, the early settlers are forced to "forego capital gains in order to prevent the entry of "undesirables" into the community" (128). And, as Buchanan explained,

"this sacrifice of capital gains on possible land holdings may be more than offset by the retention of a greater share of taxpayers' surplus" (128). The costs of creating these excluding devices could thus be too large compared to the benefits. This meant that the early settlers will not enact, and then enforce, the laws that would be necessary to exclude the potential free riders arriving in the group although it would be collectively rational to do so. They were indeed in the situation of a Samaritan's dilemma.

After having acknowledged the possibility of such dilemmas, Buchanan did not spend more time analyzing them in the 1960s. Much to the contrary. He tended to ignore them. One of the reasons can be found in the phenomenon he described in his 1959 comment: free riders are difficult to exclude once they are in the group; it is, therefore, crucial to exclude them before they enter the group. Unsurprisingly, thus, Buchanan studied and proposed mechanisms aimed at excluding potential free riders. This is what "clubs" are—"the theory of clubs is, in one sense, a theory of optimal exclusion, as well as one of inclusion" (Buchanan, 1965a, 13). Provided that property rights are correctly defined. Indeed, Buchanan wrote, property rights should "be adjusted to allow for optimal exclusion" (13). What if property rights could not be adjusted? Then, Buchanan concludes, "the " free rider" problem arises." (13) Which means that free riders would benefit from the good provided by the club members. Or, in other words, a Samaritan's dilemma would take place.

Did Buchanan acknowledge the problem? No. He had unlocked a door that he would not immediately push open. In "An Economic Theory of Clubs" (1965a), Buchanan insisted on the importance of "allowing for more flexible property arrangements and for introducing excluding devices" (14), as if he did not want to focus too much on the risks of letting free riders enter clubs. In "Ethical Rules, Expected Values, and Large Numbers" (1965b), he also stressed that the large number dilemma could be avoided by "reducing group size, or at least modifying the rules so that something similar to small-group results emerges" (10). In other words, he put the emphasis on the possibility to change property rights and re-design groups to avoid free riding and the occurrence of the Samaritan's dilemma. This is what he repeated a few years later, in *The Demand and Supply of Public Goods* (1968a), noting that in "*small* groups,… the possibility of excluding genuine non-conformists will normally be present" (87; italics in original).

At the end of the 1960s, Buchanan admitted the existence of the problem more frequently. Two papers from 1967 and 1968 are, from this perspective, particularly important. In "Cooperation and Conflict in Public-Goods Interactions" (1967), Buchanan demonstrated again how the size of groups affects the individuals' decision to cooperate and contribute optimally to the provision of a public good. Thus, in small groups, did he claim, "there will remain some motivation for the individual to behave strategically" and "some incentive for... tacit cooperation..., but as the group size grows this incentive becomes increasingly faint while the pressures toward "anti-social" behavior become increasingly strong." (115). Also, he demonstrated that strategic behaviors would produce different results depending on whether individuals choose between two or three possible strategies. With two strategies—to share the provision of the public good and to behave independently from others—,

the outcome of the game could either be the optimal cooperative solution in which both players cooperate or the worst possible situation in which no one cooperates. Both were possible, even if the second one was more probable. Then, when a third strategy was introduced—exploitation or free riding—, the outcome of the game would be unstable: "Once this additional opportunity for genuine "exploitation" of the other person is recognized, dominance no longer characterizes the utility payoff matrix." (114) He thus admitted the possibility of asymmetric outcomes as in the Samaritan's dilemma. However, and this is important to note, Buchanan built his game matrix to show that such outcomes would not last.

And that was also the point he made in "A Behavioral Theory of Pollution" (1968b), that Buchanan came even closer to describing a Samaritan's dilemma. Buchanan demonstrated that free riders in "a cooperating depollution club" could achieve a "higher utility level than that achieved by club members" (355). In other words, Buchanan had demonstrated the existence of a Samaritan's dilemma—an asymmetric situation in which some individuals cooperate and others do not. He even noted that this equilibrium was "suboptimal in the Pareto sense" (355). However, at this stage, he had not yet admitted that this situation could be a trap and that the cooperating club members could be facing a dilemma. Buchanan then noted that the cooperating members of the club could accept the situation, "acquiesce in the free-riders' gains, if they observe the latter to be reinforcing, even if feebly, their own "depollution" efforts" (355). Free riding could be tolerated if it was not too important, if free riding was not an obstacle to the provision of the public good and if free riders evidenced a disposition to change their behavior. Or, put differently, Buchanan did not (at least, seem to) believe yet that the asymmetric situation of a Samaritan's dilemma could and would last.

These articles were the last Buchanan wrote before acknowledging a problem that he had identified in the late 1950s and that was implicit in his approach to cooperation and free riding. About two years later, in 1970, he started to write "The Samaritan's dilemma", admitting that exploitation was a possible and stable (durable) outcome of an interaction between two individuals adopting a different rule of conduct.

6 Conclusion

There is no doubt that "The Samaritan's dilemma" was, as it is usually said, written in, and in reaction to, a specific context. Buchanan was pessimistic, very much affected by the situation in the Academia and, more broadly, in the American society when he wrote his essay. The latter, however, perfectly fits into the frame Buchanan had built to analyze cooperation, and conflict, in public economics. His probabilistic approach of free riding, his analysis of how the size of groups influences cooperation, does not only imply that individuals will cooperate if they are in a small group and behave strategically, or that they will not cooperate in large groups. As we have shown in this paper, Buchanan's insistence on small number environments also suggests that asymmetric and conflictual outcomes—under the form of exploitation—are possible

and indeed inevitable. It suggests the possibility of situations that can be depicted by a game like the Samaritan's dilemma.[2]

Buchanan was aware of that very early but admitted it only at the turning of the 1960s to the 1970s. Later, Buchanan will insist more frequently on the risks that such games, and their off-diagonal outcomes, could represent (for instance, Buchanan, 1977, 1993, 1998; Buchanan and Lomasky 1984). As Otto Lehto and John Meadowcroft note, Buchanan was then very much concerned that "the off-diagonal outcomes…could be relatively stable" (2021, 149). In other words, he was concerned that Samaritan dilemmas could take place. Now, Samaritan's dilemmas are much more problematic than Prisoner's dilemmas. While the latter corresponds to an anarchic situation, a situation of war of all against all that obviously requires a specific action. The former is not an exceptional situation. It corresponds to an equilibrium in a stable political-legal order that all the individuals accept, because they directly benefit from it or because it would be too costly or too difficult to get out of it.

That was one of the most important aspects and lessons of Buchanan (1975) essay. Individuals were not ready to accept the responsibilities that go with helping others—which includes not being charitable in the short term to secure long-term benefits. They do not accept, put differently, to become "better" individuals (see Buchanan, 1979, 100). This anticipated the individuals' incapacity to assume the responsibility for their actions or choices that necessarily goes with the freedom that Buchanan put forward forty years later in "Afraid to Be Free: Dependency as Desideratum" (2005).[3]

References

Boettke, P., & Marciano, A. (2020). *The soul of classical political economy. James M. Buchanan from the archives*, Mercatus Center.

Boettke, P. J., Marciano, A., & Stein, S. (2021). They never walked alone. Workshops, seminars, conferences and the making of Virginia Political Economy. *Revue d'Économie Politique, 131*(5), 729–751

Boettke, P. J., & Martin, A. (2010). Exchange, production, and Samaritan dilemmas, mimeo, http://mpra.ub.uni-muenchen.de/33199/

Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review, 40*, 289–329.

Bovard, J. (1983). Feeding everybody: How federal food programs grew and grew. *Policy Review, 26*(Fall), 42–51.

Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *American Economic Review, 90*(4), 847–868.

Dughera, S., & Marciano, A. (2022). Supervise me if you can. Relational feelings, incentive pays and supervisory violations. *Journal of Economics* (forthcoming).

[2] By contrast with what Buchanan and Roger Congleton demonstrated—namely that Samaritan dilemmas can result from majority cycling—, we have shown that the dilemma is independent from the rule used to make the political decisions.

[3] I thank John Meadowcroft for pointed this out to me.

Buchanan, J. M. (1961a). Simple majority voting, game theory and resource use. *Canadian Journal of Economics and Political Science, 27*(3), 337–348.
Buchanan, J. M. (1961b). Comments. In J. M. Buchanan (Ed.), *Public finances: Needs, sources, and utilization* (pp. 122–129), Princeton University Press.
Buchanan, J. M. (1965a). An economic theory of clubs. *Economica, 32*(125), 1–14.
Buchanan, J. M. (1965b). Ethical rules, expected values, large numbers. *Ethics, 76*(1), 1–13.
Buchanan, J. M. (1967). Cooperation and conflict in public goods interactions. *Economic Inquiry, 5*(2), 109–121.
Buchanan, J.M. (1979). Natural and Artifactual Man. *What Should Economists Do?,* Liberty Press.
Buchanan, J. M., & Lomasky, L. (1984). The matrix of contractarian justice. *Social Philosophy and Policy, 2*(1), 12–32.
Buchanan, J. M. (1999). *1968a.* The demand and supply of public goods.
Buchanan, J. M. (1968). A behavioral theory of pollution. *Economic Inquiry, 6*(5), 347–358.
Buchanan, J. M. (1975). The Samaritan's dilemma. In E. S. Phelps (Ed.), *Altruism, morality and economic theory* (pp. 71–85). Sage Foundation.
Buchanan, J. M. (1977). Law and the invisible hand. In S. H. Bernard (Ed.), *Interaction of economics and the law* (pp. 127–138). Lexington Books.
Buchanan, J. M. (1978). Markets, states, and the extent of morals. *American Economic Review, 68*(2), 364–368.
Buchanan, J. M. (1993). How can constitutions be designed so politicians who seek to serve 'public interest' can survive and prosper? *Constitutional Political Economy, 4*(1), 1–6.
Buchanan, J. M. (1998). Majoritarian logic. *Public Choice, 97*, 13–21.
Buchanan, J. M. (2005). Afraid to be free: Dependency as desideratum. *Public choice*, 124 (1/2), Policy challenges and political responses: public choice perspectives on the post-9/11 world (july), 19–31.
Buchanan, J. M., & Kafoglis, M. Z. (1963). A note on public goods supply. *American Economic Review, 53*(3), 403–414.
Buchanan, J. M., & Tullock, G. (1962). *The calculus of consent: logical foundations of constitutional democracy*. The University of Michigan Press.
Coyne, C. J. (2008). *After war: The political economy of exporting democracy*. Stanford University Press.
Fleury J. -B., & Marciano, A. (2018a). The making of a constitutionalist: James Buchanan on education. *History of Political Economy*, *50*(3), 511–548.
Fleury, J.-B., & Marciano, A. (2018b). The sound of silence: A review essay of Nancy MacLean's *Democracy in Chains: The Deep History of the Radical Right's Stealth Plan for America. Journal of Economic Literature, 56*(4), 1492–1537.
Futagami, R., Kamada, K., & Sato, T. (2004). Government transfers and the Samaritan's dilemma in the family. *Public Choice, 118*(1/2), 77–86.
Gibson, C. C., Andersson, K., Ostrom, E., & Shivakumar, S. (2005). *The Samaritan's dilemma.* Oxford University Press.
Lehto, O., & Meadowcroft, J. (2021). Welfare without rent seeking? Buchanan's demogrant proposal and the possibility of a constitutional welfare state. *Constitutional Political Economy, 32*(2), 145–164.
Marciano, A. (2013). Why market failures are not a problem: James Buchanan on Market Imperfections, Voluntary Cooperation, and Externalities. *History of Political Economy, 45*(2), 223–254.
Marciano, A. (2015). Buchanan on pro-social behaviors: Why is ethics necessary. *Œconomia*, *5*(3), 295–311.
Marciano, A. (2016). Buchanan's non-coercive economics for self-interested individuals: Ethics, small groups, and the social contract. *Journal of the History of Economic Thought, 38*(1), 1–20.
Marciano, A. (2021). James Buchanan: Clubs and alternative welfare economics. *Journal of Economic Perspectives, 35*(3), 243–256.

Musgrave, R. A. (1939). The voluntary exchange theory of public economy. *Quarterly Journal of Economics, 53*(2), 213–237.
Musgrave, R. A. (1959). *The theory of public finance*. McGraw Hill Book Co.
Olson, M. (1965). The logic of collective action. *Public Goods and the Theory of Groups*, Harvard University Press.
Phelps, E. (ed). (1975). *Altruism, morality, and economic Theory.* Russel Sage Foundation.
Rajan, R. G., & Subramanian, A. (2005). Aid and growth: What does the cross-country evidence really show?" NBER working paper w11513, https://doi.org/10.3386/w11513, http://www.nber.org/papers/w11513
Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics, 36*(4), 387–389.
Skarbek, E. C. (2016). Aid, ethics, and the Samaritan's dilemma: Strategic courage in constitutional entrepreneurship. *Journal of Institutional Economics, 12*(2), 371–393.
Stone, D. (2008). *The Samaritan's dilemma: Should government help your neighbor?* Nation Books.
Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy, 64*(5), 416–424.
Tiebout, C. M. (1961). An economic theory of fiscal decentralization. *Buchanan, 1961* (pp. 79–96). Princeton.
Wagner, R. E. (2005). Redistribution, poor relief, and the welfare state. In J. Backhaus & R. E. Wagner (Eds.), *Handbook of public finance* (pp. 385–405). Springer.
Williamson, C. (2010). Exploring the failure of foreign aid: The role of incentives and information. *Review of Austrian Economics, 23*, 17–33.

Felwine Sarr's *Afrotopia*—An Utopian Vision Fruitful for Africa?

Rigmar Osterkamp

1 Introduction

Felwine Sarr, the author of *Afrotopia*, is a Professor of economics at the University Gaston Berger in Saint-Louis, Senegal. He is known to a broader public also outside of Africa since the French President Emmanuel Macron assigned him, together with Bénédictine Savoy, the task of producing a report about the restitution of African cultural assets from Europe to Africa (2018). Sarr's *Afrotopia* was first published in French (2016), later in English (2019) and German (2020), always under the same title. In *Afrotopia*, Sarr develops a vision for future cultural, social and economic developments in Africa which would amount to a radical break with the current situation.

The aim of this article is to identify the core ideas of Sarr's vision and to describe them, as far as possible, in his own words (according to the English edition), while avoiding the redundancies the book contains. Critical comments and appraisals, particularly those concerning the usefulness and the chances for realization of Sarr's vision, are left for the end of this article.

The text is organized into six chapters which, however, do not correspond to the structure of the book. The first five chapters provide Sarr's views: a description of the problems (1.); an analysis of the causes of the problems (2.); a list of the aims of Sarr's Afro-Utopia (3.); a discussion of the necessary measures to realize his vision (4.); a speculation on the effects of the realization of the vision for the whole world (5.). The last chapter (6.) provides a critical appraisal of Sarr's *Afrotopia.*

R. Osterkamp (✉)
Bachstr. 34 B, 83673 Bichl, Bayern, Germany
e-mail: rigmar.osterkamp@gmx.de

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_20

2 Sarr's Identification of the Problems Africa is Facing

Sarr's personal and cultural background is in West Africa, particularly Senegal where he is born, though he includes other African regions in his analysis. He ordinarily speaks generally about "Africa", "Africans", "African people" or "African peoples".

Sarr proposes rather generally that a typical „African being "suffers from „narcissistic and psychological wounds that have been inflicted upon him [during colonial and later times, R.O.] and that, today, are expressed by way of lack of self-esteem, amounting to an interiorized inferiority complex". He observes "still a tendency to deem whatever may come from the West as being better". Expressions such as *White man's science*, widespread in Africa, he regards as "indications of the current self-exclusion […] from the common scientific tradition." Sarr goes even so far as to maintain there is "a pathological […] lack or absence of selfhood that translates into an inability to think for oneself, to judge and evaluate things on one's own." (p. 62 f).

He believes that this is going to change, fortunately. At least "a certain segment of the African youth […] is educated by their [own, i.e., African, R.O.] cultures **and** shaped by modern knowledge the same way as everyone else, and more specifically, (is) no longer clinging onto any kind of colonial complex [… or] part of the pathological relationship their elders maintained with respect to the former colonists and colonial powers." (p. 65 f).

Nevertheless, traditional values and behavioral norms be still vivid: One can "emphasize that in contemporary African societies despite the dilution of a tradition's capacity to regulate behavior, certain traits of traditional culture persist, notably, those tied to prestige expenditures, to investment in symbolic goods, to the injunction of generosity, and to the interiorized notion of a duty to assist." (p. 52).

This also means, Sarr maintains, that the behavior of an African cannot be described as that of a *homo oeconomicus*. The African instead behaves as a *homo africanus*, whose "decisions are to some degree motivated by logics of honor, of redistribution, of subsistence, of gifts and counter-gifts." (p. 53) The economies of traditional African societies "were marked by the fact that the production, distribution, and possession of goods were regulated by social ethics, the goal of which was to guarantee everyone's livelihood through the redistribution of resources and the right of each member of the community to receive help from the whole society if the need arose. … This flowed from a concept … of an economy at the service of the community." (p. 55).

Sarr sees these traditional economic systems being "subverted by the capitalist economy, which is largely focused on individual profit, having forgotten its original functions." (p. 55) The traditional values and behavioral norms have been displaced by what Sarr regards either as "the usual buzzwords—development, emergence, Millennium Development Goals (MDGs) […]—which […] have served merely to project the myths of the West onto the trajectories of African societies" (p. 1), or by "the keywords of the episteme of modernity in the West", namely "progress, reason, growth, order". (p. 6).

But not all African heritage is gone: "despite the various mutations under way, the cultural context of contemporary African societies […] remains a powerful determiner of the economic choices of its members." (p. 53).

3 Sarr's Cause Analysis

The central cause for Africa's economic and cultural problems is seen by Sarr in the Western concepts of "development, economic emergence, growth and struggles against poverty … [which are] exported to the world over since the fifteenth century, in favor of technological advance—via cannonballs and, when necessary, blunt objects to the skull." (p. xiii).

The process of implanting Western values, technical ideas and ideologies into African societies took place over hundreds of years. "The transatlantic slave trade (deportation) and colonialism were synonymous with the draining of wealth and resources as well as people, the destruction of societies, institutional distortions, the rape and pillaging of cultures, and finally the alienation of dominated societies into unscrupulous trajectories." (p. 34).

Besides enslavement and colonialism, Sarr mentions other causal roots for the unsatisfactory economic and social development of many African countries, namely: "poor economic management by the leaders of young, independent African nations who for the most part had made poor economic choices, and [..] power relations that were and continue to be unfavorable to the continent as much in the international arena (international rules of business) as in the choices of the strategic options with regards to economic politics (an absence of autonomy in the choice of these options: the structural adjustment programs, the consensus and post consensus from Washington, Millennium Development Goals, etc.)." (p. 36 f).

Sarr is convinced that the main causes for the discrimination Africa suffers are rooted in the colonial past and are still effective today. "Formal independences were conceded in exchange for the perpetuation of a system of political, economic, and cultural dependence, all to maintain control over the resources of the African continent. The predation of these resources still continues today through imbalanced contracts and an unequal exchange to exploit Africa's natural resources." (p. 38) Sarr even sees, in countries such as Ivory Coast, Gabon and Senegal, examples of a successful "economic re-colonization [..] by former colonial powers." (p. 38) The advent of China in Africa is regarded by him as "deleterious for the continent: a bit of infrastructure in exchange for the pillaging of Africa's natural resources and the colonizing of its land." (p. 38).

For Sarr, another main cause for Africa's problems is the university system. "By establishing the university system in Africa […] the colonial authority standardized the representation that it wanted to present of itself as well as the representation that Africans had of themselves: an image inscribed in subalternity. The colonial university thus allowed for the inscription of the other into its narration within a position of inferiority". (p. 88) It is then no wonder that also "international university

cooperation functions according to the same schema. Since the large majority of funding for scientific research in Africa comes from external sources, the scientific agenda is consequently also standardized from outside Africa. African researchers thus find themselves in a position of informers or subcontractors of questions and issues that are in no way a result of their own epistemological priorities." (p. 88).

Rhetorically, Sarr asks, "how do we provide a younger generation of Africans with a positive representation of themselves, based on theories elaborated outside of Africa whose goals are to keep them within a position of subalternity?" (p. 88).

Since Sarr is an economist, it is not surprising that the science of economics plays a prominent role in his causal analysis of problems. Strategic long-term plans, MDGs, structural adjustment obligations, and economic growth as an overriding goal—all prescribed by international financial organizations and obeyed by African countries—"are largely of the neoclassical variety and attempt, without any sort of discernment on their part, to apply these neoclassical visions to the trajectories of African countries. What's more, most of these strategic plans recommend a more improved insertion of the country's economy into the market of globalization, as if the latter were a neutral market or had the country's best interest in mind". (p. 94).

The focus on strategic planning and economic growth is, in Sarr's view, "derived from an uncritical return to the use of the concept of emergence, corresponding to the economic upturn in the five stages of economic growth according to Rostow." (p. 94; Rostow 1960).

The concept of "development" is seen by Sarr as "one of the West's entrepreneurial expressions", used worldwide to affect the "dissemination of its myths and social teleologies." (p. 4) After colonialism as a civilizing mission had been discredited, "development erected itself as an unquestionable norm of progress within human societies, […] thus denying a diversity of trajectories as well as the modalities of responses to the challenges that are posed to them." (p. 4 f).

The notions of "well-being, progress, growth, and equality are" according to Sarr, "the key concepts of Western cosmology that condition its reading of the real, a reading that it imposed onto other peoples and nations through the mytheme of development." (p. 8).

Consider, for example, the concept of Gross Domestic Product (GDP), important for measuring the level and change of economic activity. Sarr criticizes the use of GDP "to classify and place nations into a hierarchy […] placing some nations higher and some lower." (p. 2) But he also appreciates "efforts.. currently under way to enrich these measuring systems and their indicators so as to make them more nuanced and complex in order to better take into account dimensions of life outside the mere notion of added value: education, health, quality of life, the quality of the environment." (p. 3).

On a more fundamental level, however, Sarr is of the opinion that the GDP concept contains a "false evaluation of individual and social life. Life cannot be measured by the soup bowl. Life is an experience, not a performance." (p. 3).

Sarr wants to underline his view that the established science of economics is mainly inappropriate for Africa and argues the "cultural values [economists] inherit

or acquire have a profound and often unknown influence on their perceptions and attitudes [… which] influence their ability to explain the economic reality they observe, even more so when this reality is different from their own." (p. 48) His conclusion is: "Therefore it appears that the dominant economic thought conveys a culture, a unique view of the world and of man (*homo oeconomicus*). […] Economic thought develops a collection of beliefs and values that it protects and promotes." (p. 48).

Sarr criticizes the usual welfare economics due to its utilitarian understanding of societal welfare. By contrast, he appreciates the approaches of Amartya Sen ("capabilities", 2009) and Martha Nussbaum ("care economy", 2000) "because they make it possible to avoid the pitfall of economism and to reintegrate economy into a larger social system." (p. 57) Moreover, he is of the opinion that such an approach had the particular effect that "public policies would be assigned with promoting the psychosocial functioning of groups." (p. 57).

4 The Aim of Sarr's Afro-Utopia

Sarr defines his African Utopia as an "active utopia that takes as its tasks the cultivation of vast and open spaces of bountiful possibles in order to help them flourish." However, to provide sufficient room for such an active utopia to evolve, Africans must untangle their "collective imaginary" from the "dominant episteme [development, economic growth etc., R.O.]" of the West. (p. xiii f).

For attaining this aim, Sarr identifies a necessary prior step: "a philosophical, moral, and political critique of the ideology of development." The fallaciousness of this ideology lies, according to Sarr, in the "folly of the modern West [which] consists in positing its specific manner of reason as sovereign." (p. 9) Thus, it is necessary "to remove ourselves from our dependence on the rational and mechanical model that has dominated the world." (p. 10).

Such a newly won self-consciousness also entails the insight that "Africa no longer needs to try to catch up with the rest of the world. It no longer needs to run on the same paths the world indicates for it, but rather to deftly walk the path it will have chosen for itself." (p. 115) Moreover, for the insistence on a truly African way forward, Sarr sees a specific reason: "Africa's status as the oldest child of humanity means that it bears a great responsibility. It must be prudent and extract itself from following the irresponsible path of globalization that could endanger the social and natural living conditions for all of human life. [..] Africa must refrain from taking part in the childish competitions and rivalries between nations who eye each other […] to see who has accumulated the most resources and wealth, the most technological gadgets, the strongest sensations" (p. 115).

Abandoning Western paths of development is also necessary because, as Sarr believes, the technoscientific reason is "exhausted" and has led its "civilizational consequences [into] impasses". Thus, "we must heed the call for a renewal of the very sources of the imaginary and of a thought coming from an elsewhere." (p. 80).

For Africa to have a better future, Sarr sees a necessity to encourage and refresh the traditional system of a *relational economy*. In such a system, goods and services are exchanged not only on the basis of their utility. Rather, the participants form "authentic relations … independently of the market value […] outside of any truly material consideration or thought". (p. 58) Sarr even envisages the possibility of the emergence of a "collective intelligence": "This relational economy could be at the root of a collective intelligence at the heart of a community (a group, an enterprise, a peasant cooperative)". (p. 58 f).

However, the relational economy is endangered by the "classical" economy as Sarr writes with reference to the French sociologist Maurice Obadia. "According to Obadia (2012), the culture of the classical economy can negatively influence the relational economy and result, for those who measure their relational production by minimum cost, in the construction of a negative relational economy." (p. 61).

Not only the "African collective imaginary", but also the social sciences, economics included, must be, according to Sarr, "de-westernized". Attempts in that direction exist. But Sarr criticizes many such approaches and maintains: "Western forms of knowledge do not exhaust all possible methods of scientific inquiry." This leads Sarr to a rather fundamental conclusion: "we must begin an inquiry into what the real is." (p. 81).

In this sense, Sarr mentions the philosophers Mamoussé Diagne (2014) and Bonaventure Mve-Ondo (2013). Their works concern "a theory of knowledge staked at the limits of the Western vision of what a form of knowledge is. They emphasize the diversity of approaches to the real according to different civilizations and eras […] The borrowed path of Western reason is merely one among many. Mve-Ondo calls for a refusal of the exclusivity of a logocentric episteme and of the controlled verification of reason by way of the specific mode of written thought." What he envisages instead as a source of insight and knowledge is the use of traditional African myths and oral narratives. (p. 83).

5 The Necessary Measures

According to Sarr, the areas in which radical reforms are necessary are nearly all embracing: "The economic, political, and cultural spheres are pillars of the social structure that must be renovated and rebuilt. The psychological sphere is the fourth of these pillars and not the least fundamental of them." (p. 62) The first precondition for such a fundamental renovation is that African economies must begin "to operate in accordance with their own driving forces", i.e., in their own cultural contexts. The second necessary precondition is "to consider the social project in its entirety by analyzing the multiple interactions of its environmental dimensions: those that aim to ensure the conditions for existence (the economy, ecology) and those whose goal is to work towards an understanding of the meaning of existence itself (culture, philosophies, orders of purpose)." (p. 44).

Economy: The economic sphere—for Sarr meaning the economy proper, the behavioral patterns of persons engaged in production and exchange, as well as the particular thinking of economists—is, Sarr is convinced, determined by outside (Western) forces and is no longer rooted in an African socio-cultural context. But this is not easily reversed. What is necessary is "a better understanding and analysis of the determining factors of the individual and collective choices of the societies in question by applying an appropriate conceptual apparatus." (p. 53).

Such a better understanding, however, can only be attained by a rather fundamental procedure, namely by addressing "the dominant economy's deficit in the realm of theory. The analytical frameworks based solely on the premise of methodological individualism fail to take into consideration the factors that determine the economic behavior of *homo africanus*." (p. 53) "Thus a better understanding of the bases of their behaviors—their psychology, their culture, and their social realities—would make economic policies more efficient." (p. 54).

After having eliminated the theoretical deficits of the usual science of economics, a "second challenge consists in improving how we consider and reflect upon what individuals and groups target and value as their ultimate goals, across all mechanisms of social exchange, and to assign to these goals, through a principle of efficiency, the appropriate instruments for their achievement." (p. 55) Sarr assigns different tasks to the economic and cultural spheres. The latter "would be tasked with defining the goals that individuals and groups value", while "economics, as a discipline, would have the function of exploring ways to allocate resources to attain these goals." (p. 55) Sarr is convinced that by such an assignment of functions, "economics would give up its hegemonic tendency of shaping all social practices and would once again become an order of means subordinate to the ends valued by individuals and groups." (p. 55).

Sarr refers to and joins most economists when he mentions core areas of a better economic policy in Africa, e.g., increased investment in human capital and infrastructure, food security, technological innovation, and an overall improvement of productivity. In one respect, Sarr even agrees with a theorem of the predominant economic wisdom when he, together with "most economists" recommends "an increased cultivation of the comparative advantages already stemming from Africa's factorial endowments". (p. 42).

Regarding the UN project of the MDGs, Sarr employs a critical stance. On the one hand, he asks rhetorically: "Who can argue with the reduction of poverty?" (p. 93) On the other hand, the MDGs are seen ultimately to be "nothing more than teleological appropriations of these countries' futures." Moreover, after a fifteen-year period, "the masters would then verify and see if the lessons had been duly learned and applied." However, questions of national health, nutrition, education, etc., "would be things that so-called *developing* [Sarr's emphasis] nations would take care of and apply by way of their own programs, their own priorities and horizons for achieving their objectives." (p. 93).

Politics: Sarr considers a possible basis for a policy that aims at individual and social well-being. That basis "could involve defining a minimum ethical standard shared by all with agreed-upon goals—education, health, nutrition, fundamental rights—and then establishing lists of desired functionings by formulating normative

hypotheses based on a society's values as reflected by current social or religious theories." (p. 57).

To realize such an approach, it is recommended to organize a societal "debate during which individuals deliberate on the dimensions of their lives that they consider meaningful." (p. 57) Another possibility may be: "to develop a group of dimensions that represent people's values based on empirical evidence gathered by analyzing behavioral data or individual beliefs." (p. 57 f).

A much more concrete, if not very realistic, policy option is mentioned by Sarr on the last page of his book. "Africa is not obliged to hold on to the notion that a nation must remain confined to a specific territory … If it wants to, Africa can redraw the borders created in 1886." (p. 118).

Culture: How far Africa must go to overcome the influence of Western thinking becomes apparent in the words of the philosopher Valentin-Yves Mudimbe (1988), cited in the affirmative by Sarr: "For us Africans, our task will be to invest in the sciences, beginning with the social sciences and human sciences in order to grasp the tensions at play, to reexamine for our own sake the contingent findings and sites of enunciation". However, an even more fundamental effort is necessary, because we have to find out, as Mudimbe continues, "what new sense and what path to propose for ourselves and our search so that our discourses justify our singular existences engaged within a history that is itself just as singular." (p. 73).

Culture, Sarr is convinced, could play an important role in improving the African economy, because this sphere, "despite upheavals of a turbulent recent history […] remains completely intact [… it] could constitute the foundation of a more efficient economy, because it would be better connected to its cultural context." (p. 51) However, this seems to be more easily said than done. Sarr continues: "If we consider that culture is a transactional concept in constant redefinition, a dynamic notion constantly being renegotiated, the difficulty becomes identifying the distinctive and stable traits that could be considered as the framing structure of the core personality of groups and communities." (p. 52).

Even the content of school curricula and people's education must be chosen. Sarr remarks: "Beyond the simple necessity for quality education for the masses, there is a fundamental question concerning the nature of the various forms of knowledge that need to be promoted and transmitted." There is consensus about "the mastery of technical forms of knowledge with the goal of creating a more efficient society". But Sarr warns: "these forms of knowledge are not neutral and therefore imply that their impact on society must also be understood." (p. 71).

Sarr's considerations become much more concrete when he takes up the idea of Mudimbe (1988), Wiredu (1980) and Ngũgĩ wa Thiong'o (1986) and asks whether "the desired African discourse could be achieved by replacing European languages with African languages." These philosophers and writers "estimate that returning to and rediscovering the vitality of African languages would not only allow for the decolonization of minds and imaginaries but also reveal interiorities and signifying universes inscribed within an order of the world that intimately shapes meaning for Africans." (p.76).

A particular problem arises with the universities in Africa whose beginnings go back to colonial times and are, thus, not really African in Sarr's view. "In order [to attain] a fruitful enculturation and appropriation of the university, it is not simply a question of reforming the university system inherited during colonization, but rather of totally deconstructing it in order to build the foundation of a new African university responding to the demands of its societies. This deconstruction must traverse a demythification of Europe with a strategy of reconquering Africa's own being in the world." (p. 89).

The reestablishment of the universities in Africa along African lines is a rather fundamental problem because it "requires reworking the social sciences, starting with the epistemological integration concerning the objects, methods, and status of knowledge produced by the social and human sciences as they are applied to African realities." This even means: "Deconstructing colonial (ethnological) [Sarr's parentheses, R.O.] reason starts with a radical critique of the produced discourses, their theoretical frameworks, their ideological underpinnings, as well as the logic used to 'pathologize Africans' [an expression of Mudimbe, R.O.] and dominate them." (p. 74).

Sarr sees a "generation of young African thinkers" [who] have taken up this [methodological] debate and inscribed it within a dialogue of epistemes and shared experiences." (p. 78) Sarr affirmatively mentions the philosopher Nadia Yala Kisukidi (2015) whose aim is, in her own words: "creating new critical epistemologies that are neither overly disciplinized nor didactic […] and which take into account the situation of utterances of the subjects who were once colonized, dominated, not only within the order of pre-constituted forms of knowledge but also within their own open and active traditions." Kisukidi's proposal is, thus, to construct "the conditions of a new epistemology" which means, in her words, to "rethink historical experiences founded on domination in order to reform them within a shared history; a hope for reciprocal recognition, giving back to everyone their history, culture, and dignity." (p. 78 f).

This task is fortunately supported by African religions, cosmogonies and onto-mythologies which are still "primary or initial motors, whose productive force is profound and sustainable (despite the grafts they have undergone)". (p. 85 f) Indeed, "one must decipher and fortify [these forces] within an open activity of recognition of self." Sarr continues, "to do this, there must be a revolution in the recognition of various modes of knowledge formation, signaling an end to the Western intellectual hegemony through the elucidation on non-Western forms of knowledge that continue to flourish and be preserved within African societies. This revolution is not only necessary, it is urgent." (p. 86).

6 The Effects of a Realized Afro-Utopia on the Entire World

The project of truly re-erecting African culture would lead to a "new presence of Africa in the world". Sarr approvingly refers to the Senegalese writer Alioune Diop [1910–1980], who said once that this African cultural renaissance "will have as one of its effects to help increase the profundity and maturity of human consciousness." Sarr adds that this is "the only objective worthy of being assigned to it." (p. 113).

A "profound cultural revolution" is also necessary "for Africa to be a positive driving force, … so as to uplift humanity". (p. 115) "Africa must participate in the work of the edification of humanity by building a more responsible civilization, one more concerned about the environment, about a harmony and balance between different orders, of the generations to come, of the common good, and of human dignity: Africa must once again become a poetic civilization." (p. 116).

It is necessary "to exhume and revive the profound humanism of [African] cultures, such as *jom* (dignity), *téraanga* (hospitality, mutuality), *kersa* (modesty and humility), *ngor* (a sense of honor)". "Such is the profound cultural revolution that the African continent must undertake. […] From this day forward, as in the time of the first rising suns, Africa will once again become the spiritual lungs of the world." (p. 118).

Africa's cultural revolution may also have global economic effects. "The large availability of natural resources on African soil, as well as sources of renewable energies, means we can choose more responsible modes of production. And since the world is once again turning its attention toward Africa's resources—coveting them and pandering to it in order to continue the maddening direction of production—Africa has the opportunity to impose a civilizational shift by refusing to perpetuate these models of production and accumulation of wealth. By shutting off [this] engine, Africa has time to decide the destination of the locomotive." The Western countries "should learn to build and to partake in a prosperity without growth." (p. 117).

One global effect of developments in Africa on the world, independent of the success of a cultural revolution, is sure: Africa's quickly rising population will enhance Africa's political weight in the global arena. Moreover, Sarr points to another consequence of African population growth, namely for the global relevance of the French language. African youth make French not simply one African language among many, but a language that will only survive and maintain its position as an international language with the help of the French-speaking populations on the African continent." (p. 66).

7 An Appraisal

The utopian vision developed in Sarr's *Afrotopia* is first a negative one: Africa has to overcome its perceived own fixation to the West, to Western notions, theories, ways of thinking, and developmental paths. The positive vision, developed in *Afrotopia*,

consists of a proposed return to pre-colonial African values and behavioral norms. Sarr names it "a profound cultural revolution"—revolution in its literal meaning. This approach amounts, thus, more to a Re-topia than to a U-topia.

Such values and norms are, Sarr is convinced, deeply rooted in the African tradition. They have become submerged through the colonial and post-colonial contact with the West, but are still alive and can be revitalized. To such values and norms belong, to mention only some, African cosmologies, the relational economy, prestige consumption, and the oral transmission of history and myths.

While the book does not undertake to design a concrete positive utopian vision for Africa, it elaborates in much detail on the way towards such a new society. The concretization and realization of Sarr's Afro-Utopia should be attained through nothing less than scientific research by philosophers, cultural and social scientists, economists and psychologists. However, this way is long, because it must begin, Sarr is convinced, by questioning Western epistemology and Western social sciences—in the first step: Western economics—and finally by replacing these scientific approaches with African concepts.

To rethink the foundations of Western philosophy and social science in an African sense is a demanding project, albeit of course not impossible. Philosophical and social science reasoning is, contrary to, e.g., natural science and medicine, a relatively cheap affair for a society to finance. African scientists can immediately start—and many have started, as Sarr's references show—to delve into the ambitious project of enriching philosophy and social sciences with African traits.

Two questions, however, arise. The first is: would that be of any use? Do Africans really want a far-reaching "de-westernization"? Do they want to abstain from "development" in the Western sense, i.e., from a continuous improvement in the material standard of living, improved standards of health care and education included? Would ordinary African people (let alone politicians) be prepared to dispense with exports of raw materials—as Sarr stipulates in his Afro-Utopia—in order to deter the West from a destructive path?

The second question is: Will Sarr's very fundamental but also single-sided interpretation of Africa's problems eventually become a valuable contribution to mitigating these problems—or will the effect mainly consist of a gigantic self-employment project for African philosophers and epistemologists?

The question *who in Africa does really want (and understands) Sarr's cultural revolution* can also be raised with respect to his postulation that the African university system must be completely dismantled, not merely reformed, before it can be re-erected in an African sense. Sarr does not indicate how such an "African" university would look. Concrete concepts are likewise avoided when Sarr speaks about one of his favorite ideas, namely substituting the Western languages in Africa with truly African languages.

"Development", Sarr is convinced, is simply an outflow of a Western ideology which is not necessarily relevant for other parts of the world, let alone for Africa. Sarr instead favors a return to a pre-colonial Africa as a preferable alternative to "development". Yet, his questioning of "development" without specifying an alternative echoes a slogan of the former (east German) GDR. After it had become unrealistic to

"catch up and overtake" the Western standard of living, the new slogan became "overtaking without catching up"—whatever that might have meant. Sarr's alternative to development is similarly unclear.

On many pages of *Afrotopia*, Sarr claims that African countries should have their *own* programs, priorities and time horizons—instead of those prescribed by Western countries or financial institutions. This demand is of course a very basic one and part of national self-determination. Although one can hardly maintain that there are no indigenous programs and priorities in Africa, there should certainly be more. However, it is Sarr who nurtures doubts about the capability of Africans to achieve this when he speaks about the "inability to think for oneself, to judge and evaluate things on one's own." However, the reason for that inability is taken to be the West's deep embeddedness in Africa's "collective imaginary". It is only Sarr's cultural revolution that can expel Western thinking. Should one conclude that truly African programs and priorities will have to wait until Sarr's revolution has succeeded?

Sarr's view of "Western" philosophy and social science (he rarely speaks about the natural sciences or medicine) is ambivalent. On the one hand, he states—with some reason—that science is not the only approach to understanding the world. On the other hand, it is science—from epistemology to economics, though all in a de-westernized form, of course—which is the center of his approach to letting Africa find a way back to itself.

Given the many problems of African societies, it is obvious that a fundamental reorientation in Africa—of thinking and acting, of norms and values, of economy and policies—could be a reasonable undertaking. Many Africans will agree to such a project, at least in principle. However, it is highly questionable whether Sarr's specific proposal for reconstructing Africa is what many Africans would consider a useful project. They may particularly call into question the enormous detour Sarr prescribes by starting with epistemological questions before coming to improvements in everyday life. In any event, Sarr does not pose the question of whether his approach would garner broad popular acceptance in Africa. Nor does he explain who is going to fund the extensive, fundamental research seen by him as necessary for putting Africa on a better path.

An important factor in African cultural alienation is not mentioned in *Afrotopia*: religious proselytizing, first by Islamic, and later by Christian missionaries. Likewise, not mentioned is the issue of development aid, although today many African and Western social scientists and economists are convinced that poorly-designed post-colonial development aid was—through a multitude of channels—an important cause for many problems in Africa.

Contrary to what Sarr suggests, it is not only African scientists who recognize, appreciate and want to preserve "African thinking". Such a perception can also be found in the work of Western anthropologists like Lévi-Strauss (1962), who coined the notion of "pensée sauvage" in his work but without any negative connotation. Another example is the psychoanalyst Paul Parin (1963) who learned from the West African Dogon people that "the whites think too much".

Sarr's description of African problems focuses largely on cultural, psychological and intellectual aspects, while he only mentions in passing the unsatisfactory economic, social and political situation in many African countries. Recurrent famines, secessionist attempts, Islamist kidnappings, civil wars, the problem of "failed states", and nepotism are not mentioned. The high rate of population growth is touched upon, but is not considered as a problem.

Sarr's wordy and repeated description of the relational economy and the traditional African values and behavioral norms suggests something like a pre-colonial African idyll. This grasps, however, at most a very tiny part of the then reality.

Sarr is correct in pointing out the catastrophic damage inflicted upon Africa through the transatlantic slave trade. But African chiefs also have played a role, not mentioned by Sarr.

It would have been useful if Sarr had accentuated the differences between his understanding of "African thinking" and that of Western scholars. However, it may be that Sarr evades such a comparison due to his basic suspiciousness vis á vis anthropology which he reveals not least by citing Mudimbe's famous *bon mot*, that anthropologists had created the "primitives" in order to study them.

For the next edition of *Afrotopia*, Sarr could analyze how other countries, such as India, Japan, China, and Vietnam, managed to deal with foreign cultural influences in the nineteenth and twentieth centuries which must have initially appeared similarly overwhelming.

The Western reader of *Afrotopia,* in particular, will recognize that Sarr's methodical approach, putting specific emphasis on the guiding work of philosophers, epidemiologists, human and social scientists, reprises Plato's idea that philosophers should rule. The question remains, however, whether and when such a vision and the way to realize it will ever prove to be a fruitful contribution to mitigating Africa's problems.

References

Diagne, M. (2014). *Critique de la raison orale.* Les pratiques discoursives en Afrique noire.

Kisukidi, N. Y. (2015). "Décoloniser la philosophie. Ou de la philosophie comme objet anthropologique", in: *Présence Africaine.*

Lévi-Strauss, C. (1962). *La pensée sauvage.* Plon.

Mudimbe, V. -Y. (1988). *The invention of Africa. gnosis, philosophy and the order of knowledge.* Bloomington.

Mve-Ondo, B. (2013). À chacun sa raison. Raison occidentale et raison africaine, Paris.

Ngũgĩ wa Thiong'o (1986). *Decolonising the mind: The politics of language in African literature.*

Nussbaum, M. (2000). *Women and human development. The capabilities approach.* Cambridge UP.

Obadia, M. (2012). "Économie relationelle et économie matérielle", in: *Les Caheirs du Sol*, *9*.

Parin, P., Parin-Matthèy, G., & Morgenthaler, F. (1963). *Die Weißen denken zuviel. Psychoanalytische Untersuchungen bei den Dogon in Westafrika.* Atlantis, Zürich.

Rostow, W. W. (1960). *The stages of economic growth: A non-communist manifesto.* Cambridge UP.

Sarr, F. (2020). *Afrotopia, (German).* Matthes & Seitz.

Sarr, F. (2019). *Afrotopia, (English).* University of Minnesota Press.
Sarr, F. (2016). *Afrotopia, (French).* Éditions Philippe Rey.
Sarr, F., & Savoy, B. (2018). *Rapport sur la restitution du patrimoine culturel africain.*
Sen, A. (2009). *The idea of justice.* Harvard UP.
Wiredu, K. (1980). *Philosophy and an African Culture.* Cambridge UP.

When Census is an Election: A Game-Theoretic Analysis of Over-Reporting of Headcount

Vikas Kumar

> In a severely divided society … an election can become an ethnic head count … a census needs to be 'won.'
> — Horowitz (2000: 196)

1 Introduction

The revisional round of the twelfth decennial population census of India was completed on 05 March 1991. Four months later the economy was liberalised. So, the first post-liberalisation census was conducted in 2001. Liberalisation affected government statistics in three ways. First, it altered the structure of the underlying political economy rendering post-colonial statistical systems designed in the 1950s less effective under the changed circumstances. The government belatedly addressed this problem through piecemeal implementation of the recommendations of the National Statistical Commission (2001). Second, the downsizing of the government due to structural adjustment following economic liberalisation limited the budget of statistical agencies and constrained their ability to hire staff and modernise operations, which affected the quality of sample surveys as well as administrative statistics. In the case of the latter, the government belatedly tried to address some of the problems through schemes such as Support for Statistical Strengthening. The budget constraint of the 1990s does not seem to have affected the 2001 Census, though, as it was conducted a decade after liberalisation when government finances had improved.

Prof. Manfred Holler has been very supportive of the larger body of my work at the interface of state and statistics of which this chapter is a part. In fact, this chapter draws upon a paper presented at the Adam Smith Seminar. It therefore gives me great pleasure to contribute to this Festschrift in honour of Prof. Holler. I am grateful to an anonymous referee for comments on a draft of this chapter. The usual disclaimer applies.

V. Kumar (✉)
School of Development, Azim Premji University, Bengaluru, India
e-mail: vikas.kumar@apu.edu.in

M. A. Leroch and F. Rupp (eds.), *Power and Responsibility*,
https://doi.org/10.1007/978-3-031-23015-8_21

In fact, the budget allocation for census increased fourfold from ₹3 billion in 1991 to ₹12 billion in 2001 (Banthia, 2001). Third, the retreat of the state intensified the race to capture public resources in economically backward regions that were bypassed by the economic transformation triggered by liberalisation. The degeneration of nationalist conflicts after the end of the Cold War also intensified sub-national conflicts over redistribution in India's ethno-geographic periphery. These developments affected census that is one of the most important inputs to the sharing of public resources.[1] The fact that the electoral constituencies were going to be delimited in 2002 after three decades[2] intensified the conflicts over census statistics. In several "peripheral" states these conflicts spilled over to census that is used to delimit constituencies. The use of flawed census data in delimitation was challenged in courts. As a result, delimitation had to be postponed in several states until the first census taken after 2026. So, until 2031 the distribution of seats in these states will be governed by the 1971 Census. In some of these states, the data for 2001 revealed abnormal changes with respect to earlier censuses (Fig. 1).

Nagaland, a small northeastern state of India, reported an abnormally high decadal growth rate of more than 64 per cent in the 2001 Census (Fig. 2). In neighbouring Manipur, similarly high growth rates were reported in several subdivisions in the hill districts (Fig. 3). Jammu and Kashmir, a large northwestern state, reported an increase in the population of Kashmir Valley that altered the longstanding trend of the relative population shares of the regions of the state (Fig. 4). In each of these cases, the reported changes cannot be accounted for by demographic factors. In the 2011 Census, Nagaland reported a negative decadal growth rate, the first such instance in post-independence India. The high growth rates, however, persisted in Kashmir Valley and a few hill subdivisions of Manipur, which also reported sharp changes in the child population. The headcounts of these three states suffer from large, non-random coverage and content errors, while statistics on child population, urbanisation, migration and population density also reveal abrupt and counterintuitive changes (Agrawal & Kumar, 2013, 2020a, 2020b and Kumar, 2023). Similar problems have been reported to a lesser extent in other states as well. However, the growing political polarisation has meant that successive governments have failed to address the communalisation/politicisation of census statistics.

[1] Census is referred to in the following articles of the Constitution that govern the delimitation of electoral constituencies of various kinds: Articles 55 (Manner of Election of the President), 81 (Composition of the House of the People), and 170 (Composition of the Legislative Assemblies). Later amendments added references to census in Articles 82 (Readjustment after each census), 243 (The Panchayats [village local bodies]), and 243P (Municipal bodies). Articles 330 (Reservation of seats for Scheduled Castes and Scheduled Tribes in the House of the People), and 332 (Reservation of seats for Scheduled Castes and Scheduled Tribes in the Legislative Assemblies of the States), which govern affirmative action, too, refer to census. See also Agrawal and Kumar (2020a: 192–247, 291–296) for a discussion of constitutional and policy linkages between allocation of public resources and government statistics in India.

[2] Until the early 1970s, state assembly and parliamentary constituencies were delimited after each census. This practice was discontinued in 1976 to ensure that the states that were relatively successful in reducing fertility did not suffer a loss of seats in the parliament (Kumar, 2022a).

Fig. 1 Jammu and Kashmir, Manipur and Nagaland. *Note* **a** Map is not to scale and may not represent authentic international borders. **b** The borders of Jammu and Kashmir correspond to the status of the territory during 2001 and 2011 Census. The state was bifurcated into two union territories in 2019. *Source* Prepared by author using a map template available at https://mapchart.net/india.html

The director of census operations for the 2011 Census in Nagaland attributed the abnormalities in the state's headcount to the fact that people often viewed census through a political-economic lens. She observed:

> Many equated it [Census] with electoral rolls and saw the decadal Census exercise as an opportunity to increase the population in villages and towns to increase the vote bank.... These problems were also compounded by the Developmental model followed in the State in

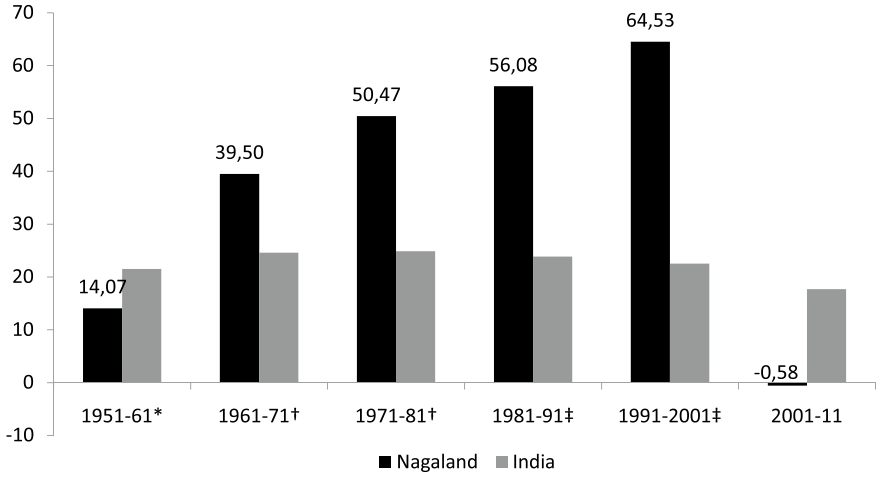


Fig. 2 Decadal Population Growth Rates (in per cent), Nagaland and India, 1951–2011. *Notes* (i) * The growth rates for the period 1951–61 are computed after excluding Tuensang, which was not fully enumerated in the 1951 Census. (ii) † The growth rates have also been adjusted for the period 1961–81 due to the change in the reference date in 1971. (iii) ‡ Identifies years in which the growth rate of Nagaland was the highest among all the states of India. *Source* Primary Census Abstracts of the respective years

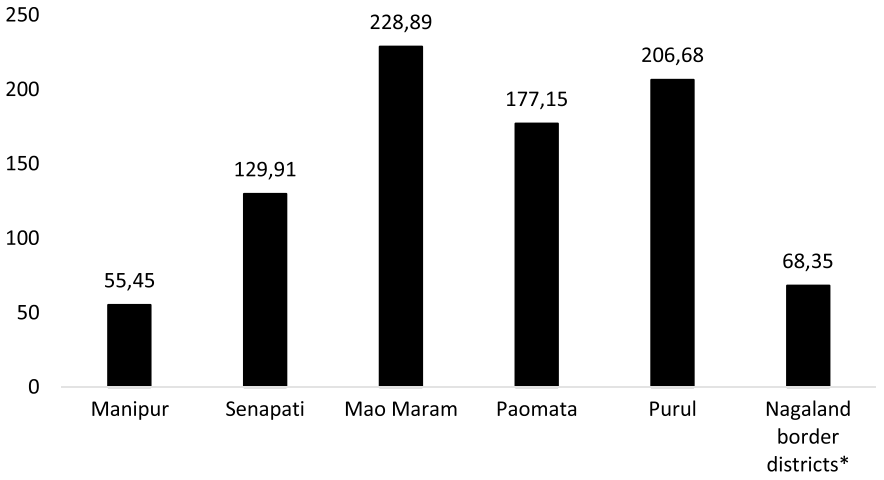


Fig. 3 Population Growth Rates (in per cent), Manipur, 1991–2011. *Notes* (i) The growth rate is the percentage change in population between 1991 and 2011, i.e., two decades. (ii) Mao-Maram, Paomata and Purul are subdivisions of Senapati district of Manipur. (iii) * Identifies the three districts of Nagaland—Kohima, Phek and Peren—which share a border with Senapati. *Source* Agrawal and Kumar (2020b)

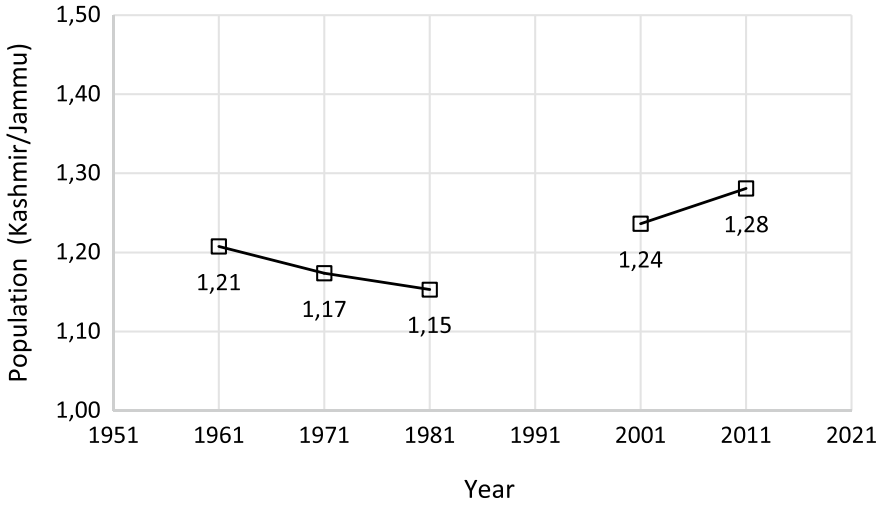


Fig. 4 Relative Population Share of Kashmir and Jammu, 1961–2011. *Notes* (i) Census was not conducted in Jammu and Kashmir in 1951 and 1991. (ii) Kashmir and Jammu refer to administrative divisions of the erstwhile state of Jammu and Kashmir. *Sources* Primary Census Abstracts of the respective years

> which allocation of funds to Village Development Boards is made on the basis of population and households in a village. This naturally led many to try and increase the fund flow into their villages by showing non-existing population and households in the Census records. (GoI, 2011: viii, 2)

Her counterpart in Jammu and Kashmir, who served as the director of census operations for the 2001 Census, added that '[e]veryone says neighbouring areas exaggerated their population but their area own was under-enumerated' (Feroze Ahmed, *Interview*, Jammu, 04 December 2019). While it is true that political and economic competition between localities and communities affects government statistics, this in itself is not sufficient to explain why people manipulate census that does not directly govern individual or household entitlements. There are three reasons why people themselves over-report headcount or support the mobilisation in this regard by ethno-statistical entrepreneurs. First, the public outreach of census inadvertently emphasises its usefulness for *individual* welfare (Kumar, 2023). Second, the same set of officials enumerate residents of an area during census, enroll citizens as voters in electoral rolls and identify the poor through household surveys. Individuals often first inflate numbers in the last two and inflate their headcount during census enumeration to maintain consistency across databases. Agrawal and Kumar (2020a: 230) show that if the electoral roll of an area is inflated in a given year, the following census, too, overcounts people. Third, when the government examines the reliability of aggregate data on electorate or welfare schemes, it often relies upon consistency checks using census data (see, for instance, ECI, 2017).

Agrawal and Kumar (2020a) show that competitive manipulation has affected the quality of census data in India's social and geographical peripheries. In particular, they show that widespread over-reporting in the 2001 Census of Nagaland was driven by political and economic factors such as the distance from nearest town, distance from bus routes, distance from plains, share of rural population, degree of ethno-linguistic fractionalisation, voting rate, demand for the formation of new ethnically homogenous districts and whether electoral rolls were inflated in polls conducted before census. Competitive manipulation has also vitiated headcounts in other states of India and several other developing countries.

Ethno-political contests undermine the provision of public goods, where 'the government monopolises access to basic goods and services valued by a majority of the population, and in which government officials have individualised discretion over how these basic goods and services are distributed' (Chandra, 2009). Under these circumstances, 'voters decide between politicians, not by assessing their policy positions, but by assessing whether a candidate will favour them in the distribution of patronage' (ibid). Politicians, too, face a 'trilemma'—'the responsibility to be true to his office, to help members of his clan, village, and tribe individually, and to bring forth development for his constituency as a whole' (Tinyi & Nienu, 2018: 174).

In such settings, 'ethnic numbers game[s]' are fuelled by the 'close association between the desire for group hegemony and the democratic ethos of universal suffrage and majority rule' that motivates 'groups in competition… to adjust their numerical ratios for sectional hegemonic interests' (Stephen Olugbemi quoted in Bookman, 2013: 20–21). In his survey of ethnic conflicts in developing countries, Horowitz draws attention to the entanglement of elections and manipulation of census:

> As an entitlement issue, the census is a splendid example of the blending of group anxiety with political domination…. Disputes over census results in ethnically divided societies are common…. In a severely divided society … an election can become an ethnic head count. Now it is clear that a census needs to be 'won.' So the election is a census, and the census is an election. (Horowitz, 2000: 194–196)

So, while the design of census assumes that individuals are standalone respondents, people *qua* respondents believe that they are embedded in a context of interactive decision-making. In fact, communal anxiety is often exacerbated by the design of census operations. In India, the design tries to minimise the possibilities of double counting by simultaneously counting people across the country through an extended de facto (synchronous) method of enumeration. However, this engenders distrust in ethnically divided states where the impartiality of the government is suspect. Under such circumstances communities may try to secure their future entitlements or protect their existing entitlements by boosting their numbers if they fear that the government will be unable, or unwilling, to stop other communities from manipulating statistics.

The mismatch between the design of censuses and popular perception of the exercise has serious consequences for the quality of census data but has not received sufficient scholarly attention. The literature on census has discussed the possibility of competitive manipulation in different contexts but has not formally examined the

structure of the strategic decision-making involved in such situations.[3] This chapter presents a game-theoretic analysis of the problem. Kumar (2020) discusses this using normal-form games for two communities. The present chapter adds further structure to the problem by explicitly accounting for the size of public pie, cost of manipulation and punishment for manipulation and generalises the setting to *n*-communities.

The discussion in the chapter is organised as follows. Section 2 draws upon Kumar (2020) to illuminate the structure of the strategic setting with the help of normal-form games involving two communities. This is followed by a model for *n*-communities in Sect. 3, which shows that, among other things, the degree of over-reporting by a community shares a non-linear relationship with the total number of communities. Further, larger communities are less likely to over-report their population, but a community is more likely to over-report its population if it is surrounded by larger communities. Finally, Sect. 4 offers concluding remarks.

2 The Setting

Communities and administrative divisions (such as villages, sub-districts and districts) are the units of analysis in our case as the share of seats in the legislature and development funds, which are formally linked to government statistics including headcount, are at stake and these accrue to the community as a whole rather than to any individual. A large part of development funds are allocated for building and maintaining public infrastructure and supporting public services that can be accessed by everyone in the community. We are not including government officials as a player as they do not have the means to enforce their will against communities manipulating en masse and even if they report the problem follow-up is not guaranteed unless the political leadership is interested in intervening.[4] Even if the leadership is willing

[3] Discussions on the Indian census have focused on its legal foundations (Kumar 2020), public outreach (Kumar, 2020, 2022b), delays in the release of data (Agrawal & Kumar 2020b), quality of metadata (Kumar, 2021), coverage errors (Agrawal & Kumar 2013, 2020a; Guilmoto & Rajan, 2013, Bhat, 2018, Kumar, 2023) and content errors in data on tribes (Agrawal & Kumar 2020a; Guha, 2003; Kulkarni, 1991; Verma, 2013), castes (Bhagat, 2006; Chaudhury, 2012), religious groups (Gill, 2007) and language groups (Agrawal and Kumar 2020a; Brass, 1974). In addition, there are comparisons of census and other databases such as the National Sample Surveys (Kasturi, 2015), discussions on the quality of census data *qua* inputs to statistical (Guilmoto & Rajan 2013) and policy (Agrawal & Kumar, 2020a) analyses, and discussion on the interface between censuses and maps (Agrawal & Kumar 2017, 2020a) and census and sample surveys (Agrawal & Kumar 2014, 2020a). Many of these contributions discuss issues that involve different kinds of interactive decision makers but hardly any tries to explore the strategic behaviour of actors. The literature on census in other developing countries reveals a similar gap (Agrawal & Kumar, 2020a; Kumar, 2023).

[4] The Census Act, 1948 tries to balance diverse interests and gives limited powers to government enumerators. See Kumar (2020, 2023) for an analysis of the severe limitations of legal and administrative measures to prevent manipulation of demographic statistics and Agrawal and Kumar (2020a) for a discussion of the views of government officials and politicians in this regard.

to address the problem, corrective action is undertaken only in the next round of enumeration (Kumar, 2020, 2023).[5]

For the sake of simplicity, we will assume that there is no migration. Otherwise, communities can possibly increase their population by assimilating immigrants, who in turn will have to choose whether they want to be assimilated. Including migration will introduce a new type of player and complicate the analysis. A related assumption ensures that the administrative boundaries of districts are coterminous with community boundaries, i.e., one district is home to one community, or the minority community of each district is evenly distributed in such a way that they cannot influence electoral outcomes. This assumption holds good in large parts of India's social and geographical peripheries (see, for instance, Agrawal & Kumar, 2020a: 212). If we introduce geographical concentration of minorities within districts that will complicate the analysis by generating incentives for gerrymandering to deny them a share in legislative power and creating room for alliance between co-ethnic groups across districts.

Further, we will assume that communities/administrative units move simultaneously, i.e., they are not aware of others' choices before they choose. Most countries follow the extended synchronous method of enumeration that involves counting residents simultaneously across the territory during a specified period. No community controls all the nodes of bureaucracy and political leadership to be able to access real time data for other communities when enumeration is underway. In India, even senior bureaucrats and the minister-in-charge get to know aggregate figures about a week to ten days after the reference date and the community wise breakups are available much later. In other words, it is very difficult for any community to access reliable information about the reported headcount of others during the course of enumeration. In fact, until the detailed tables are published one can only guess the choices of other communities *after* the provisional results of census are published. Therefore, during enumeration, each community must choose without knowing if and by how much other communities have over-reported their numbers. The information asymmetry triggers acute communal anxiety because census is mostly a decennial exercise and any loss due to over-reporting by others can be redressed only after a decade if not later. For instance, communities of Nagaland that believe that they were allocated fewer seats in the 1970s than warranted by their population share had to wait until 2001 to seek a redressal.

We further assume that communities *qua* players have complete information. This is because under the assumption of no migration we can argue that the communities have shared space over a long period. Finally, we assume that the interaction of communities can be modelled with the help of a one-shot game because census

[5] In order to include bureaucrats in the game, we will have to shift to a dynamic setting and introduce two types of government officials—field enumerators (junior officials) and decision-makers (senior officials)—and deal with the possibility of moral hazard. This will complicate the analysis whose primary objective is to highlight competitive manipulation of census engendered by strategic interactions.

is a decennial exercise, and every round of census is not (directly) linked to the distribution of public resources.[6]

We have so far suggested that in the context of decennial census communities *qua* strategic competitors are locked in a one-shot simultaneous move game of complete information. We need to add another assumption, namely, over-reporting of headcount is *pre-emptive* in nature insofar as it is meant to check unfavourable changes in allocation in the future (*after* the publication of census data). In absence of this assumption, we will have to work in a dynamic setting with at least two periods and introduce additional players such as government officials.[7] We will also assume that if both communities choose to inflate their numbers, their relative population share does not change. While not unreasonable, this last assumption is not necessary for the analysis and will be relaxed in the next section.[8]

The central feature of the interaction during census is as follows. A community is worse off if it does not over-report its population while other communities do so. There are two reasons behind this. First, the community loses share in legislative power and public expenditure that are apportioned according to population shares. Second, a fresh census is unlikely if other communities have over-reported their headcounts as the new status quo would be favourable to them. However, over-reporting is costly as people have to be mobilised to overstate their household size when enumerators visit their house and government officials need to be convinced, bribed or coerced to accept manipulated figures rather than alert the higher authorities. The risk of detection and punishment by authorities is not the only reason why manipulation is costly. There is also a risk of aggravated conflict with competing communities denied their share of legislative constituencies and development funds due to flawed data. After the 2001 Census, several communities approached courts with their grievances related to the use of flawed data for the delimitation of constituencies. It is important to note that when all the communities overstate their size, each one of them expends resources to maintain its share in population even as the risks of conflict with rivals and authorities are not mitigated and development planning is impaired due to non-availability of reliable statistics.

---

[6] Until recently, the 1971 Census was the most important determinant of federal redistribution in India. After recent revisions, a combination of 1971 and 2011 censuses is used in formulae for federal redistribution.

[7] *Altruistic* manipulation involves intervention 'with an eye on the distant future, say, the next generation,' while *defensive* manipulation is meant to 'ensure that the latest census figures do not contradict manipulated figures reported in the last census or manipulated figures reported to access welfare schemes during the intercensal period' and *prospective* manipulation aims 'to secure a favourable policy change, for example, the creation of a new administrative unit' (Agrawal & Kumar, 2020a: 193).

[8] In Nagaland, the pervasive manipulation of headcount in 2001 more or less preserved the distribution of population based on 1991 Census (Agrawal & Kumar, 2020a: 227–228). In Nigeria (Fawehinmi, 2018) and Pakistan (Khan, 1998: 481; Weiss, 1999: 687, 691), successive censuses have reproduced historical population shares of provinces. In an extreme case, Lebanon has not conducted a census since 1932 to insulate its fragile power sharing arrangement from demographic changes (Horowitz, 2000: 195).

In the present setting, each community has two pure strategies – *Over-report* and *Not Over-report*. The outcomes of the interaction are governed by the following conditions.

(a) A community is better off over-reporting its headcount if others also do so, i.e., (Over-report, Over-report) $\succ$ (Not Over-report, Over-report).
(b) (Over-report, Over-report) $\not\succ$ (Not Over-report, Not Over-report) because effort invested in manipulating statistics goes waste as the status quo is maintained.

Following the first two conditions, (Over-report, Over-report) is neither the best outcome nor the worst outcome for either community.

(c) (Not Over-report, Not Over-report) is not the worst outcome for either community as they do not incur the cost of manipulation under the status quo. If, however, this does not hold good, communities would perversely prefer simultaneous manipulation even though this entails costs without concomitant gains.
(d) (Over-report, Not Over-report) $\succ$ (Over-report, Over-report), i.e., a community prefers to over-report when the other community is not over-reporting.
(e) The relative ordering of (Not Over-report, Not Over-report) and (Over-report, Not Over-report) depends on the context, e.g., the effectiveness of the government in detecting and prosecuting over-reporting.

Two different orderings of the outcomes, which are related to two different types of players, are possible under the above conditions.

Type I: (Not Over-report, Not Over-report) $\succ$ (Over-report, Not Over-report) $\succ$ (Over-report, Over-report) $\succ$ (Not Over-report, Over-report).

Type II: (Over-report, Not Over-report) $\succ$ (Not Over-report, Not Over-report) $\succ$ (Over-report, Over-report) $\succ$ (Not Over-report, Over-report).

The above orderings can support three different games depending on how communities rank the outcomes (Fig. 5). The first two games pair players of the same type (I–I and II–II), whereas the last game pairs different types (I–II or II–I).

Game 1 (Type I–I): This resembles games of coordination and has two pure-strategy Nash equilibria. Under one neither community over-reports, whereas under the other both over-report. The latter equilibrium is Pareto-dominated. There is a mixed-strategy equilibrium that can be interpreted in two ways. We can either argue that a community randomises over time between *over-report* and *not over-report* or communities comprise of two types of individuals who differ vis-à-vis willingness to manipulate figures. The mixed-strategy equilibrium is Pareto-dominated by (Not Over-report, Not Over-report).

Game 2 (Type II–II): This game resembles *Prisoner's Dilemma* and supports only one pure-strategy equilibrium under which both communities over-report (Over-report, Over-report). But this equilibrium is Pareto-dominated by (Not Over-report, Not Over-report). The equilibria of the third game that pairs different types (Type I-Type II/Type II-Type I) resemble Game 2.

The literature suggests that the threat of punishment cannot mitigate over-reporting engendered by communal distrust and weak public institutions. So, we

I

A \ B	Over-report	Not Over-report
Over-report	$\underline{(2,2)}$	$(3,1)$
Not Over-report	$(1,3)$	$\underline{(4,4)}$

II

A \ B	Over-report	Not Over-report
Over-report	$\underline{(2,2)}$	$(4,1)$
Not Over-report	$(1,4)$	$(3,3)$

I − II

A \ B	Over-report	Not Over-report
Over-report	$\underline{(2,2)}$	$(3,1)$
Not Over-report	$(1,4)$	$(4,3)$

where $4 \succ 3 \succ 2 \succ 1$ and the pure-strategy equilibria are underlined.

Fig. 5 Manipulation Games

can argue that Game 1 better represents the interaction in the context of census as it allows communities to switch away from the Pareto inferior equilibrium without any threat of punishment. In contrast, in Game 2 (and Game 3) players can move to the Pareto-dominant outcome only under the threat of punishment.

The (Over-report, Over-report) equilibrium of *Game 1* captures the pervasive over-reporting across Nagaland in 2001, when almost all communities overstated their headcount. On the other hand, the (Not Over-report, Not Over-report) equilibrium of *Game 1* captures the outcome of the 2011 Census of Nagaland when almost all communities refrained from manipulating figures. Growth rates were very high across Nagaland in 2001 and very low in 2011 (Fig. 2). As discussed later, this switch between equilibria happened due to public dialogue and awareness campaigns. The mixed-strategy equilibrium of *Game 1* relates to (a) the 1991 Census of Nagaland, where only a few communities over-reported their numbers; (b) the 2001 and 2011 Census of Manipur, where only a few communities in the northern hill districts over-reported their populations; and (c) the 2001 and 2011 Census of Jammu and Kashmir, where over-reporting was confined to Kashmir region.

3 Census Games

We will model over-reporting of headcount in a jurisdiction with T communities taking into account gains and costs of over-reporting. Communities choose the scale of over-reporting of population, which can potentially have one or more of the following effects: (a) the overall population of the state increases that in turn increases inter-state federal transfers, (b) the community that over-reports its population gains seats through delimitation of electoral constituencies within the state, and (c) the community that over-reports its population receives more development funds allocated to/raised by the state. We will abstract from the possibility of greater federal allocation (a) that will expand the size of the public pie in the state because population is only one of the criteria for federal allocations. Moreover, lagged population data are used for federal redistribution and parliamentary constituencies to avoid demographic competition between states. In other words, in India, the federal fund allocation is independent of the *current* population and, by implication, is not affected by manipulation. Note that the size of internal funds is also invariant to the degree of manipulation because a state's internal resources depend on the *actual* population. In other words, the overall size of the public pie is exogenously determined and is not affected by the degree of manipulation of the headcount. We will also abstract from the possibility that the second effect (b) could reinforce the third effect (c) in winner-take-all majoritarian democracies such as India or that smaller communities could ally against a larger community.

In the following analysis, the total size of public funds of the state and total number of seats in the state legislative assembly are denoted by D and L.[9] Further, $L_i = p_i L/P$ is the share of community i in the legislative assembly and $D_i = p_i D/P$ is the share in development funds, where p_i is the community's true population and $i \in \{1, 2 \ldots T\}$. $P = \sum_i p_i$ is the true overall population. $C_i(n_i)$ is the cost of over-reporting incurred by community i to mobilise its members and to bribe or force officials to ignore n_i ghost individuals wrongly reported along with the community's true population, p_i. The marginal cost of over-reporting varies across communities but is assumed to be constant for a given community, say, i: $C_i^{'}(n_i) = m_i$, where m_i depends on the community's political, economic, and bureaucratic power relative to other communities. We assume that F is the fine for one unit of over-reporting, whereas in practice the actual fine would be lesser.[10] Further, $1 \geq r_i \geq 0$ is the probability of getting fined and is a measure of the government's administrative capacity and/or judiciary's effectiveness, which varies across communities.

If community i over-reports its population by n_i, its gain from over-reporting can be expressed as follows, where $n = \sum_i n_i$:

$$\pi_i = (L + D)\frac{p_i + n_i}{P + n} - (L + D)\frac{p_i}{P} - m_i n_i - r_i n_i F \tag{1}$$

The first order condition for community i is given by:

$$(L + D)\frac{(P + n) - (p_i + n_i)}{(P + n)^2} = m_i + r_i F \Rightarrow (L + D)\frac{(P + n)_{-i}}{(P + n)^2} = m_i + r_i F \tag{2}$$

It can be shown that the second order condition is also satisfied. Now summing up (2) over all the T communities we obtain the following, where $M = \sum_i m_i$:

$$(L + D)\frac{(T - 1)(P + n)}{(P + n)^2} = M + \sum r_i F \Rightarrow (L + D)\frac{(T - 1)}{(P + n)} = M + \sum r_i F \tag{3}$$

By comparing (2) and (3) we can arrive at the following expression for the reported population (including over-reporting) of community i as a fraction of the overall reported population.

[9] With some additional assumptions our conclusions will hold if D depicts government spending as a fraction of the size of economy.

[10] The census legislations of most countries specify fines for individuals manipulating information without explicitly linking the fine to the level of manipulation or factoring in the possibility of collusive manipulation (Kumar, 2020). Consider two cases, one in which each household inflates its headcount by one member and another in which it inflates by two. In most countries, the total fine will be same in both the cases.

$$\frac{(P+n)_{-i}}{(P+n)} = 1 - \frac{p_i + n_i}{P+n} = (T-1)\frac{m_i + r_i F}{M + \sum r_i F}$$

$$\Rightarrow \frac{p_i + n_i}{P+n} = \frac{(M - m_i(T-1)) + \left(\sum r_i F - r_i F(T-1)\right)}{M + \sum r_i F} \tag{4}$$

Using (3) and (4), over-reporting chosen by community i can be expressed as:

$$n_i = (L+D)(T-1)\frac{(M - m_i(T-1)) + \left(\sum r_i - r_i(T-1)\right)F}{\left(M + \sum r_i F\right)^2} - p_i \tag{5}$$

By comparing the corresponding expressions of (4) for communities i and j we obtain the following relationship between the over-reporting of two communities:

$$n_j = \frac{\left(M - m_j(T-1)\right) + \left(\sum r_j - r_j(T-1)\right)F}{(M - m_i(T-1)) + \left(\sum r_i - r_i(T-1)\right)F}(p_i + n_i) - p_j \tag{6}$$

Equation (6) can be recast as follows where θ is a measure of the cost of manipulation of community j relative to other communities.

$$n_j = \theta n_i + \left(\theta p_i - p_j\right), \text{where}\, \theta = \frac{m_i + r_i F}{m_j + r_j F} \text{ for } T = 2 \tag{7}$$

We can now state the key claim.

Claim 1: If the expected fine for manipulation is very small $\left(\sum r_i F \ll M\right)$ and the mean marginal cost of manipulation is less than the per capita share of public resources ($\overline{m} = M/T < (L+D)/P$), the over-reporting ($n$) of overall population (p) in a multi-community society ($T \geq 2$) is positive and independent of fine (F). The over-reporting of community i's population is positive if its marginal cost of over-reporting is not too large compared to the mean marginal cost of manipulation for the communities, $m_i < \overline{m}T/(T-1)$

The expression for over-reporting (5) simplifies if the probability of punishment is very small ($r_i \to 0$) and/or if the marginal cost of over-reporting is much more than the fine ($F \ll M$), i.e., the expected fine is small $\left(\sum r_i F \ll M\right)$:

$$n_i|_{\sum r_i F \ll M} = (L+D)(T-1)\frac{M - m_i(T-1)}{M^2} - p_i \tag{8}$$

$n_i > 0$ holds if the following is satisfied:

$$1 < \left[(L+D)(T-1)\frac{(M - m_i(T-1))}{M^2 p_i}\right] \tag{9}$$

$$0 < \frac{M}{T-1} < \frac{L+D}{p_i}\left(1 - \frac{m_i(T-1)}{M}\right) = \frac{L+D}{P} \cdot \frac{P}{p_i} \cdot \left(1 - \frac{m_i(T-1)}{M}\right) \tag{9'}$$

$$0 < \frac{M}{T-1} < \frac{L+D}{P} \cdot \frac{P}{p_i} \cdot \left(1 - \frac{m_i(T-1)}{\bar{m}T}\right) \tag{9''}$$

Summing (9) over T communities, $N = \sum n_i > 0$ if the following expression holds good:

$$1 < (L+D)\frac{(T-1)}{MP} \Rightarrow 0 < \frac{M}{T} < \frac{M}{T-1} < \frac{L+D}{P} \tag{10}$$

Under the assumptions specified in Claim 1, the degree of over-reporting is non-zero and independent of the level of fine. The significance of the condition on marginal cost, probability of punishment (r_i), and magnitude of fine (F) in the above claim would bear elaboration. The condition suggests that a community over-reports its population when the gain in the form of legislative seats and development funds is more than the marginal cost of over-reporting. The size of public pie is larger than the cost of capturing it in societies where the public sector dominates the economy and the effective penalty for violating census law is low.

The probability of punishment (r_i) is, indeed, very small. Over the past seven decades, there have been hardly any cases in India of penal action for giving incorrect responses. Kumar (2020, 2023) points out that this is true of other democracies as well. The probability of punishment is particularly low in cases of mass violation of (census) law, which is the case dealt with here.[11] Furthermore, the level of fine F is very small in most countries (Fig. 6). So, the assumption that $r_i \ll 0$ and $F \ll M$ is justified.[12]

The following, also illustrated in Fig. 7, are entailed by Claim 1.

Corollary 1: Over-reporting (n_i) by a community i increases with development funds (D) and seats in the legislature (L). (Fig. 7a and b).

Other things being equal, an expanding public pie accentuates competition (measured by the level of manipulation/over-reporting) for public resources because communities are concerned about their relative shares rather than absolute share.

Corollary 2: Over-reporting (n_i) by a community i initially increases as the number of communities (T) increases and decreases after $T \geq \overline{T_i} = \frac{M+2m_i}{2m_i}$ (Fig. 7c).

[11] Cases of communal rioting, which involve mass violation of the law, too, are marked by abysmal conviction rates. Wilkinson (2004: 90) notes that 'prosecution and arrest rates for communal violence throughout the world have always been relatively low, and therefore it is doubtful if calculations about the marginal likelihood of prosecution affect the individual rioter very much.'

[12] Fines for violating census laws are very small across the world, but the rate of detection of violations is higher in countries with robust institutions. It is a different matter that as discussed in the concluding remarks the authorities refrain from prosecuting individuals. Note that high rate of detection but negligible rate of prosecution combined with low fines translate into very light penalties.

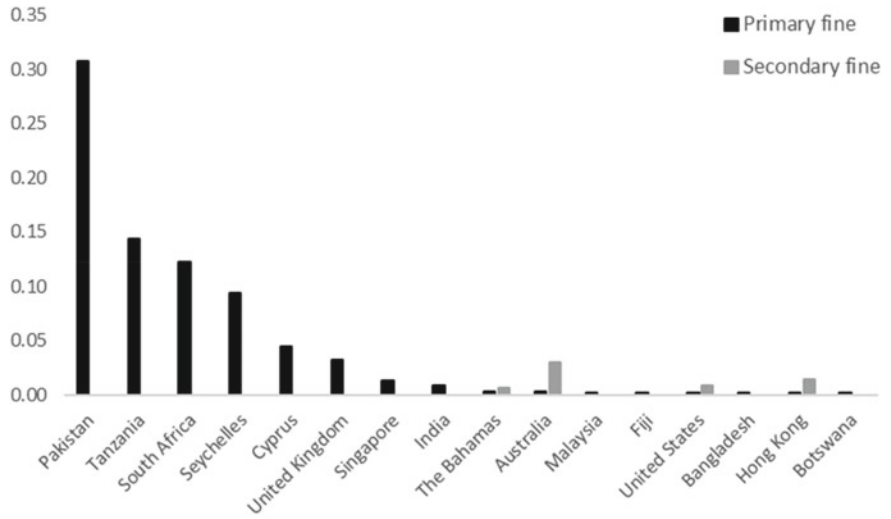


Fig. 6 Census fines relative to per capita GDP. *Note* The graph shows the ratio of fines prescribed in census legislation to GDP per capita, both in current local currency units. In countries where there is a hierarchy of fines, both the primary (offence: refusing to supply information and, in some countries, supplying incorrect information) and secondary (offence: in some countries, supplying incorrect information) fines are shown. *Sources* Kumar (2020) Fines: Census/statistics laws downloaded from the webpages of the respective national statistical/legal departments or the United Nations Statistics Division ("Laws and Acts on vital statistics system," https://unstats.un.org/unsd/vitalstatkb/KnowledgebaseCategory14.aspx). GDP per capita (current Local Currency Unit, 2017): World Bank Open Data (https://data.worldbank.org/)

The expression for $\overline{T_i}$ follows from Eq. (8). Competition is intense when there are only a few communities. As the number of communities increases, the competition decreases as the conflict dyads multiply and there is no clear competitor or clear winner and loser. This is akin to moving from polarisation, when there are a few communities, to fractionalisation, when there are many communities. Aggregate over-reporting is zero for one community. The over-reporting of community i is zero for both one community and for a very large number of communities and is highest when there are $\overline{T_i}$ communities. Agrawal and Kumar (2020a: 230) show that over-reporting decreases as the effective number of communities or ethnic fractionalisation increases.

Corollary 3: Over-reporting (n_i) by a community i (a) decreases as its own population (p_i) increases, (b) increases as other communities' population (p_j) increases, and (c) increases with other communities' over-reporting (n_j) (Fig. 7d, g, and h).

So, larger communities have lesser incentive to over-report their population. However, when surrounded by larger communities, a community has greater incentive to over-report its population. Analysing the data for 78 sub-district units of Nagaland, Agrawal and Kumar (2020a: 228, 230, 246) shows that in 2001 over-reporting decreased with increase in the size of sub-district, tribe or language group. At the

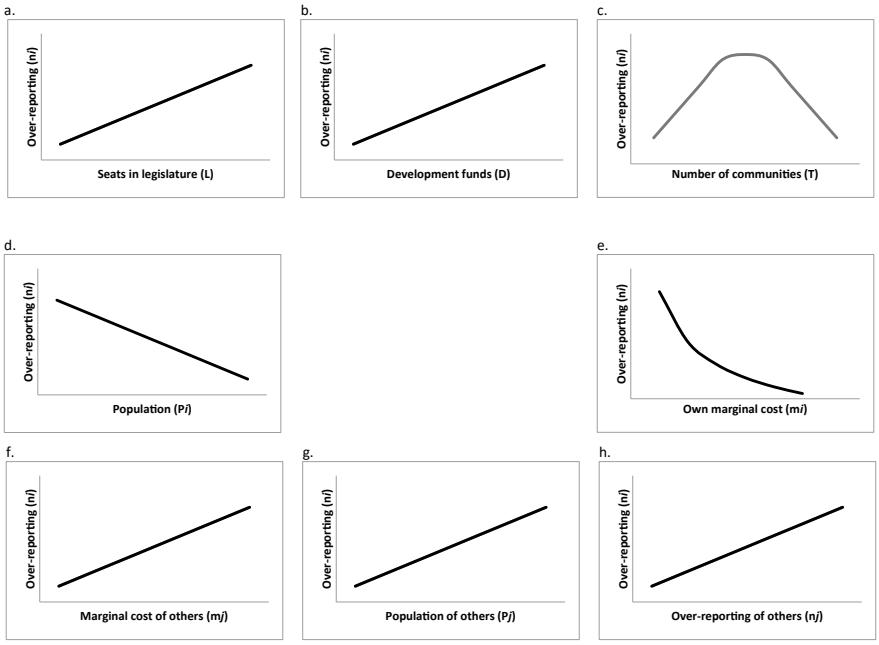


Fig. 7 The effect of different factors on over-reporting. *Source* Author

same time the demonstration effect of over-reporting by neighbouring communities was strong (ibid: 225).

Corollary 4: Over-reporting (n_i) by a community i (a) decreases with increase in its marginal cost (m_i) and (b) decreases/increases as its cost of over-reporting (m_i) increases/decreases relative to other communities (m_j). (Fig. 7e and f).

Numerically comparable communities with different costs of over-reporting, over-report their populations to different extents. If communities face identical costs, then smaller communities will over-report more relative to larger communities. The level of over-reporting can exceed the total population if the size of public pie is sufficiently larger than the cost of over-reporting. Parts of Nagaland had, indeed, over-reported their population by more than 100 per cent in 2001 (Agrawal & Kumar, 2020a: 167, 170, 172, 223).

4 Concluding Remarks

We presented a game-theoretic model of manipulation of census by communities competing over legislative power and development funding. The analysis for two communities using normal-form games suggested that there are multiple equilibria

including one in which there is no over-reporting and another in which all communities over-report their headcounts. As a result, two identical societies can report very different behaviour due to idiosyncratic factors. Further analysis explicitly accounting for gains and costs of manipulation suggested that over-reporting is most intense when there are only a few communities. Moreover, larger communities are less likely to over-report their population. On the other hand, a community is more likely to over-report its population if it is surrounded by larger communities. We argued that given the very low rates of penalties stipulated in census laws and a deep reluctance to invoke the punitive provisions, the gains from manipulation might outweigh costs making manipulation likely in societies with weak institutions and low levels of inter-community trust. The key results of the game-theoretic analysis agree with the suggestion of Agrawal and Kumar (2020a) that the quality of data is often intertwined with the nature of democratic institutions and the level of development. In other words, *data*, *development* and *democracy* deficits are intertwined.

Given the fact that these games support multiple equilibria, the level of mutual trust and trust in public institutions determines if communities manipulate numbers. Governments cannot rely on punitive provisions to deter communities from manipulation. Modern censuses assume a cooperative relation between people *qua* respondents and the government *qua* enumerator. Reliance on punitive provisions to address the problem of non-cooperation will generate a 'backlash against the census that would further depress cooperation' (Prewitt, 2003: 15, see also Prewitt, 2010: 239). Discussions with census officials and political leaders of Nagaland, where the overcount of 2001 was largely eliminated in 2011, suggested that the government did not contemplate the use of punitive measures against individuals or communities at any stage. The shift from pervasive over-reporting in 2001 to a consensus against over-reporting in 2011 was also not facilitated by any change in the punitive provisions of the Census Act, 1948 that was last amended in 1994. The Nagaland government adopted a three-pronged approach. First, it mechanically invoked the threat of punishment in general warnings. It also announced that the results of enumeration would be cross-checked with the help of church membership records (GoN, 2009) and biometric data (Assam Tribune, 2011), but neither was part of the census calendar. Second, the government took steps to improve vigilance of the census operations, assured greater transparency in census operations and facilitated confidence-building through public outreach. An advertisement campaign (e.g., "My future must be built on the truth–Correct Census means strong future!") was supplemented by political leaders (Rio, 2011: 73–74) and the civil society (Rutsa, 2010), who appealed to the Christian-moral values of people. Third, the chief minister assured the people that their *existing* welfare entitlements will not be reduced if their population contracted but future entitlements will not increase if the headcount is fraudulently inflated. This undercut the incentives for defensive as well as prospective manipulation (see footnote 8).

Last but not the least, a public meeting two years before enumeration (GoN, 2009) brought *all* political and non-political stakeholders on a common platform to build a (legally non-binding) consensus in favour of conducting a clean census in 2011. The platform allowed otherwise competing communities to arrive at a shared

understanding of the problem based on an appreciation of the importance of an accurate census to public policy. The platform resolved to support the state initiative to hold a clean census and agreed that community volunteers would help if census enumerators faced resistance in the field.

A clarification about the nature of the aforesaid interaction between communities from the perspective of games discussed in this chapter is in order. The public interaction organised by the state government took place two years *before* enumeration, which does not amount to communication *during* the play of the game. In fact, the door for strategic departure from the *non-binding* consensus remained open. The extended (synchronous) method of counting meant that as before communities were in the dark about the choices of other communities during enumeration.

However, the significance of open-ended consultations in such situations is evident from the divergent experiences of Nagaland, on the one hand, and Manipur and erstwhile Jammu and Kashmir, on the other. These states faced similar problems vis-a-vis census and delimitation, but they handled the problem very differently. Nagaland resorted to public dialogue without blaming any community. In Manipur, the Imphal Valley blamed the Naga-dominated hill districts of the state and the government tried in vain to address the problem through coercive administrative mechanisms. In erstwhile Jammu and Kashmir, the state government did not even acknowledge the problem and the dominant sections of the media and intelligentsia, too, did not discuss the issue. As a result, over-reporting persisted in the 2011 Census in both Manipur and Jammu and Kashmir. In light of this, a final point requires clarification from the perspective of game theory. Following Basu (2015) we can say that in Nagaland the government did not create a new equilibrium. Rather it helped highlight a (preexisting) equilibrium in a game of multiple equilibria by creating another focal point.

References

Agrawal, A., & Kumar, V. (2013). Nagaland's Demographic Somersault. *Economic and Political Weekly, 48*(39), 69–74.

Agrawal, A., & Kumar, V. (2014). Infirmities in NSSO Data for Nagaland. *Economic and Political Weekly, 49*(12), 20–22.

Agrawal, A., & Kumar, V. (2017). Cartographic Conflicts within a Union: Finding Land for Nagaland in India. *Political Geography, 61*, 123–147.

Agrawal, A., & Kumar, V. (2020a). *Numbers in India's Periphery: The Political Economy of Government Statistics*. Cambridge University Press.

Agrawal, A., & Kumar, V. (2020b). Manipur's Population Conundrum. *Economic and Political Weekly, 55*(48), 48–56.

Assam Tribune. (2011). PC keeps mum on population growth. 02 April. Retrieved 20 June 2019 from http://www.assamtribune.com/scripts/mdetails.asp?id=apr0211/at09.

Banthia, J. K. (2001). Mobilising Support for India's Census-Constraints and Challenges. Paper presented at UNFPA/PARIS21 International Expert Group Meeting: Censuses, Pretoria. 26–29 November.

Basu, K. (2015). "The Republic of Beliefs: A New Approach to 'Law and Economics'." *World Bank Policy Research Working Paper* No. 7259.
Bhagat, R. B. (2006). Census and caste enumeration: British legacy and contemporary practice in India. *Genus, 62*(2), 119–134.
Bhat, B. A. (2018). Reality Behind Declining Child Sex Ratio in Jammu and Kashmir. In *Emerging issues in maternal, Neonatal and Child Health in India a compendium of studies conducted by the population research centres 2016–17* (pp. 169–186). New Delhi: Ministry of Health and Family Welfare Government of India.
Brass, P. R. (1974). *Language, religion and politics in North India*. Cambridge University Press.
Bookman, M. Z. (2013). *The demographic struggle for power: The political economy of demographic engineering in the modern world*. Routledge.
Chandra, K. (2009). Caste in our social imagination. Seminar 601 (September).
Chaudhury, P. (2012). Political economy of caste in Northern India, 1901–1931. Working paper. New Delhi: Centre for Economic Studies and Planning, Jawaharlal Nehru University.
ECI/Election Commission of India. (2017). *Electoral Statistics Pocket Book 2017*. New Delhi.
Fawehinmi, F. (2018). The story of how Nigeria's census figures became weaponized. Quartz Africa, 06 March.
Gill, M. S. (2007). Politics of population census data in India. *Economic and Political Weekly, 42*(3), 241–249.
GoI/Government of India. (2011). Census of India 2011, Provisional Population Totals, Paper 1 of 2011, Series 14 Nagaland. Directorate of Census Operations; Kohima.
GoN/Government of Nagaland. (2009). *Resolutions adopted in the Consultative Meeting on Census–2011 held at the Zonal Council Hall, Kohima, on 30th September, 2009*. (Copy obtained from the Director of Census Operations, Nagaland).
Guha, S. (2003). The Politics of Identity and Enumeration in India c. 1600–1990. *Comparative Studies in Society and History, 45*, 1, 148–167.
Guilmoto, C. Z., & Irudaya Rajan, S. (2013). Fertility at the district level in India: Lessons from the 2011 census. *Economic and Political Weekly, 48*(23), 59–70.
Horowitz, D. L. (2000). *Ethnic groups in conflict* (2nd ed.). University of California Press.
Kasturi, K. (2015). Comparing census and NSS data on employment and unemployment. *Economic and Political Weekly, 50*(22), 16–19.
Khan, A. H. (1998). 1998 census: The results and implications. *The Pakistan Development Review, 37*(4), 481–489.
Kulkarni, S. (1991). Distortion of census data on scheduled tribes. *Economic and Political Weekly, 26*(5), 205–208.
Kumar, V. (2020). Census laws and the quality of census data: The limits of punitive legislation. *Statistical Journal of the International Association for Official Statistics, 36*(4), 1143–1160.
Kumar, V. (2021). Usability of India's government statistics. *Statistical Journal of the International Association for Official Statistics,* 429–445.
Kumar, V. (2022a). Delimitation, In J. J. P. Wouters & T. B. Subba (Eds.), *The Routledge Companion to Northeast India*, Routledge, pp. 115-120.
Kumar, V. (2022b). The public outreach of the Census of India and data quality. *Manuscript*.
Kumar, V. (2023). *Numbers as Political Allies: The Census in Jammu and Kashmir*, Cambridge University Press (forthcoming).
Prewitt, K. (2003). Politics and science in census taking. New York, Washington, DC: Russell Sage Foundation, Population Reference Bureau.
Prewitt, K. (2010). The U.S. decennial census: Politics and political science. *Annual Review of Political Science, 13*, 237–254.
Rio, N. (2011). *Speeches of Neiphiu Rio, Chief Minister, Nagaland, 2010*. Directorate of Information and Public Relations, Government of Nagaland.
Rutsa, X. (2010). Nagaland census 2011 begins. *Eastern Panorama*, 05 July. Retrieved 20 June 2019 from http://www.easternpanorama.in/index.php/component/content/article/59-2010/july/1065-nagaland-census-2011-begins

Tinyi, V., & Nienu, C. (2018). Making sense of corruption in Nagaland: A culturalist perspective. In J. J. P. Wouters & Z. Tunyi (Eds.),*Democracy in Nagaland: Tribes, traditions, tensions* (pp. 159–180). Kohima: The Highlander Books

Verma, A. K. (2013). Tribal "Annihilation" and "Upsurge" in Uttar Pradesh. *Economic and Political Weekly, 48*(51), 52–59.

Weiss, A. M. (1999). Much ado about counting: The conflict over holding a census in Pakistan. *Asian Survey, 39*(4), 679–693.

Wilkinson, S. I. (2004). *Votes and violence: Electoral competition and ethnic riots in India.* Cambridge University Press.