Wolfgang Buchholz · Dirk Rübbelke
Editors

The Theory of Externalities and Public Goods

Essays in Memory of Richard C. Cornes

Springer

The Theory of Externalities and Public Goods

Wolfgang Buchholz • Dirk Rübbelke
Editors

The Theory of Externalities and Public Goods

Essays in Memory of Richard C. Cornes

Springer

Editors
Wolfgang Buchholz
Institute of Economics and Econometrics
University of Regensburg
Regensburg, Germany

Dirk Rübbelke
Chair for Economics Especially Resource Economics
TU Bergakademie Freiberg
Freiberg, Germany

ISBN 978-3-319-49441-8 ISBN 978-3-319-49442-5 (eBook)
DOI 10.1007/978-3-319-49442-5

Library of Congress Control Number: 2017935980

Printed on acid-free paper

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

Acknowledgements

We thank Alison Clugston Cornes, Todd Sandler and Roger Hartley for their very valuable suggestions and comments concerning this book. Furthermore, we are grateful to Claudia Baldauf for her help in organising the review process as well as to Sebastian Otte for his support in compiling the papers. Finally, we thank the anonymous referees who reviewed the contributions to this book.

Contents

The Theory of Externalities and Public Goods: The Lifework of Richard Cornes

Wolfgang Buchholz and Dirk Rübbelke

Several months ago, after Richard had passed away, we felt that there was a need for activities honoring Richard and his lifework. In the beginning, we thought hard about the right way to recognize this exceptional researcher, colleague and friend. One appropriate option was to bring together some of those colleagues that he was most associated with and to ask them to contribute a research paper to a book honoring him. The idea was immediately supported by Roger Hartley and Todd Sandler who offered helpful advice, e.g. on whom we may ask to contribute. Of special importance to us was that Alison, Richard's wife, was also delighted with the book project.

Now, with the contributions in hand, preparing the introductory chapter to this collection of papers and, especially, providing adequate tribute to Richard's outstanding scientific work is both difficult and easy.

On the one hand, it is difficult to provide an all-inclusive appreciation of Richard's contributions to economics because his publications cover a rich variety of microeconomic aspects—ranging from general game-theoretical approaches, the theory of contests and duality theory[1] to their real-world applications, e.g., in environmental and family economics, public choice and fiscal federalism. Needless to say that Richard's many contributions on these diverse topics have been published in the most highly regarded journals.

[1] Richard's stimulating book "Duality and Modern Economics" from 1992 helped many economists to refine their techniques to depict and to solve economic problems.

W. Buchholz
Institute of Economics and Econometrics, University of Regensburg, Regensburg, Germany

D. Rübbelke (✉)
Chair for Economics Especially Resource Economics, TU Bergakademie Freiberg, Freiberg, Germany
e-mail: dirk.ruebbelke@vwl.tu-freiberg.de

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_1

Yet, on the other hand, assessment of Richard's achievements is also easy since a common thread runs through his entire work. Almost since the beginning of his scientific career, Richard has dealt with non-cooperative behavior in economies with reciprocal externalities and the approaches to overcome the ensuing inefficiencies. In the 1980s, Richard (together with Ted Bergstrom and Todd Sandler as his co-authors) and other famous economists (e.g., Larry Blume and Hal Varian) were the founding fathers of the theory of voluntary (or "private") provision of public goods. The monograph *The Theory of Externalities, Public Goods, and Club Goods*, which he published with Todd Sandler in 1986, has been a milestone in this field that broke new grounds and immediately attracted attention of the scientific community. On the cover sheet of its considerably enlarged second edition in 1996, the book was praised by the later Nobel prize winner Elinor Ostrom as "the authoritative work on public goods that all political economists doing theoretical and empirical work should have on their shelves. A major accomplishment".

Wolfgang Buchholz, as the older of the two editors of this volume, remembers quite well when he first had a look into the book some thirty years ago—and was soon fascinated by its topic and its masterly style, both regarding its fine English phrasing and its precise, but easily comprehensible formal arguments. It was this book, which motivated Wolfgang to direct his research interests to public good theory.

Since the release of this book, Richard has contributed to public good theory again and again, e.g. by exploring the partially paradoxical effects on public good supply and welfare, which result in various extensions of the standard public good model. In this context he, in particular, has considered different and changing public good production technologies, unconditional and conditional ("matching") transfers between agents, and the impure public good model where individual contributions generate private co-benefits.

With the Aggregative Game Approach (jointly conceived with Roger Hartley), Richard developed a novel tool for the analysis of Nash equilibria in voluntary public good provision. This approach advances the investigation of comparative statics of Nash equilibria substantially and, importantly, renders possible the derivation of results when there are more than two asymmetric agents. This new analytical method thus helped to overcome the "curse of dimensionality" in public good theory. As Richard wrote with regard to the Aggregative Game Approach[2]: "Aggregative games can be analysed using an approach that avoids the proliferation of dimensions as the number of players grows, and that does not rely on adopting such restrictive further assumptions as the notion of the 'representative player' or the restriction to two players".

By coincidence, or a quirk of history, voluntary provision of public goods turned into an important empirical issue in the last decades: The provision of global public goods (in particular, climate protection and disease eradication), for

[2]This is from a first draft of a book with the working title "Modelling Aggregative Games" that Richard prepared—jointly with Roger Hartley—early in the preceding decade.

which no central authority with coercive power exists so that autonomous states must act voluntarily, is considered to be an outstanding challenge for international politics. Thus, it is not surprising that the theory of voluntary public good provision as pioneered by Richard and co-authors has become an indispensable basis of the thriving field of international environmental economics. Richard has set his mark also here. Just recently, he published an article in the prestigious journal, *Environmental and Resource Economics*, in which the many applications of his Aggregative Game Approach to environmental economics are presented. It is very sad and deplorable that we have to recognize that this has been one of Richard's last publications. There is no doubt that a lot of Richard's insightful contributions will stay in the memory of our science—but it is a big loss that no new ones can be added.

Beyond, we lost a friend whose ideas inspired our general thinking and ideas. During a dinner, Richard told Dirk Rübbelke, the younger editor of this book, that he plans to publish—jointly with Roger Hartley—a book about aggregative games and remarked that a very initial draft was recently lost when his laptop was stolen. Of course, Richard was at first unhappy that he had to start writing the text again. Yet, he pointed out that he realized that the theft had (also) a good effect: He liked the 'second' version much more than the first.

Richard frequently saw opportunities where others would have seen a loss or a harm. And his curiosity, power of observation and rational attitude brought about new points of view to us. So, when Brian Schmidt (like Richard, a professor at the ANU) won the Nobel Prize in Physics in 2011 "for the discovery of the accelerating expansion of the Universe through observations of distant supernovae" (https://www.nobelprize.org/nobel_prizes/physics/laureates/2011/schmidt-facts.html), Richard mentioned that he would like to ask Brian Schmidt whether he thinks that he would also have been awarded the prize if he had found out that the expansion is not accelerating but decelerating, thus just confirming what most cosmologists had supposed before. Thereby, Richard stressed the role that fortune plays in our lives (without having the intention to play down the achievements of the Nobel Prize winner)—like the fortune that we had when we met Richard and had the chance to benefit from his research and thoughts.

In this book, we gather research papers that were influenced by and benefited from Richard's research. The first papers in this book are applying the Aggregative Game Approach, as it was of utmost significance for Richard's recent research activity.

In his contribution, *Roger Hartley* studies collective contests in which contestants lobby as groups. A key result is that group-lobbying effort is efficiently produced in equilibrium. Furthermore, Hartley points out that through the application of a decomposition theorem, the analysis of equilibria can be substantially simplified and demonstrates that—under standard assumptions—a collective contest can be analyzed by reducing it to a conventional Tullock contest between groups. This procedure allows transferring results from standard contest theory to collective contests.

Alex Dickson's contribution establishes a framework for considering 'multiple aggregate games' and for capturing settings in which, on the one hand, there is a collective element to individual actions within groups, and, on the other hand, there are externalities between groups' aggregate actions. Dickson also discusses applications of the theory in order to show its usefulness for analysing strategic interactions involving individuals in groups.

Wolfgang Buchholz and *Michael Eichenseer* analyse strategic international coalition building in global public good provision, which is an issue of particular importance in the field of climate change economics. In so doing, they develop a two-stage game where in the first stage the members of two groups of countries decide on coalition formation and in a second stage choose their public good contribution cooperatively within their respective group while the groups act non-cooperatively against each other.

In a semi-aggregative presentation of a game, *Alex Possajennikov* models individual players that have a conjecture about the reaction of the personalized aggregate to a change in the considered agent's own strategy. Evolution is supposed to select conjectures that lead to a higher payoff in the equilibrium. Possajennikov shows that for any conjectures of the other agents, only conjectures that are consistent, in the sense of being equal to the slope of the actual reaction function, can be evolutionarily stable. It is further shown that in public good games and contests, consistent conjectures are actually evolutionarily stable.

Ngo Van Long's model is also concerned with voluntary public good provision. He considers situations where contributors to a public good belong to two distinct behavioral types: Kantian and Nashian. The analysis ascertains how the expected level of aggregate contribution to the public good may change across mixed strategy equilibria when the proportion of Kantian in the population changes.

Avinash Dixit and *Simon Levin* consider the role that pro-social preferences may play in overcoming undersupply of public goods. As they argue, societies can benefit by instilling such preferences in their members and investigate an intergenerational education process for this.

In a partly similar vein, *Simon Vicary* addresses David Hume's concept of justice as a social convention and its relation to public goods. In doing so, he uses ideas expressed by Hume to formalise the idea of a social convention.

Turning to public good provision at different jurisdictional levels *Emilson Silva* examines the efficiency of decentralized leadership in federal settings where selfish regional governments provide regional and federal public goods while the benevolent central government implements interregional earmarked and income transfers.

For the investigation of voluntary public good provision in an intertemporal framework *Christian Haslbeck* and *Wolfgang Peters* use an overlapping-generations model, in which they study the conflict between young and old generations over sharing the costs of public good provision. In their setting, an intergenerational income redistribution via public debt proves to be 'neutral' with respect to public good supply. Thus, Barro neutrality (as known from the theory of public debt) meets Warr neutrality (as known from the theory of voluntary public good provision).

Todd Sandler applies public good theory to the highly topical policy field of counterterrorism. As Sandler points out, counterterrorism actions may possess both opposing and re-enforcing externalities, which lead to strategic substitutes and complements between country-specific defensive measures. In the analysis of these externalities, he draws on different concepts elaborated together with Richard, like the joint production/impure public good model.

Indraneel Dasgupta analyses the consequences of integrating large minorities into an economy with a population majority of another language. He develops a two-community model where such assimilation generates overall social gains, but has adverse distributional consequences. Total resource wastage due to ethnic conflicts may increase as well. Dasgupta's analysis in particular helps to explain why attempts to integrate large minorities into majority ethno-linguistic conventions may be associated with strong resistance although there are potential gains from such integration.

Thomas Eichner and *Rüdiger Pethig* examine carbon emissions control in a group of countries to explore the distributional incidence of mixed policies. These policies consist of an emissions trading scheme on the one hand and of national emissions taxes that overlap with the emissions trading scheme on the other. The effects of distinct policy designs on the distribution of welfare are investigated.

Karen Pittel and *Dirk Rübbelke* develop a dynamic joint-production model in order to analyse the implications of local and global pollution when two different types of abatement activities can be undertaken. In their two-country endogenous growth model, one type of abatement reduces solely local pollution while the other mitigates global pollution as well. Pittel and Rübbelke focus on the effects that the degree to which global externalities are internalised in the countries' environmental policies has on pollution and economic development, and they derive policy rules adapted to the different scenarios.

In their contribution to this book, *Bouwe Dijkstra* and *Patrick Graichen* analyse a referendum campaign as a case study of a contest. The referendum held in a small German town in the Black Forest led to the replacement of the conventional electricity supplier by a firm founded by the local environmentalists. Dijkstra and Graichen discuss both qualitative as well as quantitative aspects of the environmentalists' victory showing some relationship between public good provision and public choice theory.

References

Cornes, R. (1992). *Duality and modern economics*. Cambridge: Cambridge University Press.

Cornes, R. (2016). Aggregative environmental games. *Environmental and Resource Economics, 63*(2), 339–365.

Cornes, R., & Sandler, T. (1986/1996). *The theory of externalities, public goods, and club goods* (1st ed./2nd ed.). Cambridge: Cambridge University Press.

Efficiency in Contests Between Groups

Roger Hartley

1 Introduction

In recent years a number of articles have investigated contests between groups of players. Most of this literature is characterized by a structure in which players choose a level of input and the inputs of all members of a group determines (positively) the lobbying effort for the group through an impact or production function. The probability that a group wins the contest is the ratio of its lobbying effort to aggregate lobbying effort. The prize awarded to the winning group might be regarded by that group as a within-group public good or a private good which is to be divided according to a sharing rule to which members of the group commit before the contest is run. It might even exhibit a mixture of these characteristics, but in all cases the fundamental difference from a simple contest is the strategic tension between a contestant's incentive to increase input in order to increase the probability of winning the inter-group contest and the incentive to free ride on the input of other members of the same group. The positive externalities implicit in the production function may lead one to anticipate that the production of lobbying effort in equilibrium is inefficient. Indeed, such inefficiency, often labelled as free riding, is typically invoked to explain apparently counter-intuitive comparative statics. However, one of our main observations is that, in equilibrium, each group's lobbying effort is efficiently produced. This does not preclude a member of a group choosing zero input but this will occur if and only if equilibrium lobbying effort can be efficiently produced with no input from that member. As well as challenging conventional explanations of the properties of equilibria, this observation also leads to a two-stage procedure for studying collective contests that is considerably simpler and perhaps more insightful than working directly from first-order conditions.

R. Hartley (✉)
School of Social Sciences, University of Manchester, Oxford Road, Manchester M13 9PL, UK
e-mail: roger.hartley@manchester.ac.uk

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_2

Some of the earliest literature on contests played by groups arose from an attempt to construct a formal framework in which to study Olsen's (1965) analysis of the relative lobbying effectiveness of small and large groups. For example, Esteban and Ray (2001) conclude that, in certain circumstances, members of large groups may fare less will than those in smaller groups and ascribe this to free-riding. However, we argue instead that, in a situation where large groups perform less well than small groups, this is not due to inefficient production of lobbying effort, but rather to the fact that the prize is divided amongst more members of the group.

An early formal analysis of existence and uniqueness of equilibrium was carried out by Katz et al. (1990) who considered two groups in which the group's production function is just a sum of the inputs of members of the group and costs are linear. Baik (1993) generalized this to more than two groups but still supposed that the probability of winning was influenced by the inputs of members of groups only through aggregate input. Baik later gave a more complete development (Baik 2008), and considered the impact of budget restrictions on players. A similar model was analyzed by Riaz et al. (1995) and Dijkstra (1998) and, recently, Ryvkin (2011) has used the model to examine how lobbying effort is affected by how players are sorted into groups. However, the assumption that production functions are a simple sum of inputs and costs are linear leads to equilibria in which, typically,[1] at most one player in each group is active (chooses positive input). Recent articles by Epstein and Mealam (2009) and by Kolmar and Rommeswinkel (2013) avoid this perhaps implausible outcome by considering production functions which do not simply aggregate inputs within groups. Our decomposition procedure considerably simplifies the analysis of equilibria in such cases and allows us to demonstrate that, under plausible and standard conditions, an equilibrium exists and the profile of equilibrium lobbying effort is unique. Three recent contributions which do not satisfy these condition are by Chowdhury et al. (2013) who study group contests in which only the highest effort in a group affects the probability of that group winning and by Baik et al. (2001) and Topolyan (2014) who assume the group with the highest output wins.

When the prize is wholly or partially a private good as opposed to a pure within-group public good, adding players to a group may reduce payoffs. The clearest case is when the prize is shared equally between members of the winning group. An early study of the implications of this by Nitzan (1991) looks at the effects of different sharing rules and Ueda (2002) examined oligopolization (groups becoming inactive). Esteban and Ray (2001), draw on insights from these models in their formal examination of Olsen's famous group size paradox. Once again, the assumption made in these articles that production functions are a sum of inputs and costs are linear lead, typically, to equilibria with at most one active player in each group. In recent contributions, Nitzan and Ueda (2011, 2014) use techniques developed by Cornes and Hartley (2005, 2007) to study collective contests in which

[1] When more than one member of a group is active, the equilibrium profile of inputs within the group is not unique.

many of these restrictions are relaxed. In both of these articles, Nitzan and Ueda assume the prize is a mixture of a public and a private good and, in Nitzan and Ueda (2014) use this approach to study the relative effectiveness of more or less homogeneous groups as measured by the values members ascribe to winning. In Nitzan and Ueda (2011), they assume that sharing rules are exogenously determined but that members of one group cannot observe the sharing rule chosen by other groups. Such incomplete information takes us outside the scope of the analysis below.

In Sects. 2 and 3, we set out the model and describe the decomposition theorem. Its use allows us to avoid a direct attack on the study of equilibria of collective contests. Such an approach typically involves studying first order conditions for best responses and this leads to a system of non-linear inequalities whose complicated nature has led many scholars to simplify the analysis by imposing symmetry and/or restricting the number of groups to two. However, this may be an unwelcome restriction as analysis of contests played by individuals shows that new features may emerge in general asymmetric contests with many players (see, for example, Cornes and Hartley 2005). Our aim in this article is to show how the decomposition theorem permits us to study contests played by groups with minimal restrictions on contest success and cost functions and an arbitrary number of players. The decomposition theorem entails defining a collective cost function for each group. This is the minimal normalized aggregate cost of producing a given level of group effort and its calculation is a straightforward optimization problem. The overall contest can then be studied by finding equilibria of a conventional contest played by groups with a simple lottery-type contest success function and the collective cost functions for each group. The equilibrium efforts of each group in this reduced contest can then be used to determine full equilibrium strategy profiles of the original contest.

In Sect. 4, we study the implications of this result for existence and uniqueness of equilibria and discuss properties of equilibria, in particular efficiency issues within groups. In Sect. 5, we present a number of examples of group cost functions. In particular, we look first at linear impact and cost functions and then at homogeneous impact functions and constant elasticity cost functions. These latter assumptions include all cases of the general model we have encountered in the literature. In Sect. 6, we study rent dissipation and note that in a fully symmetric case, rent dissipation is independent of the size of groups. Section 7 applies the decomposition theorem to comparative statics and, in particular the effect on winning probabilities and payoffs of adding a new group and of adding members to an existing group. A final section concludes.

In this preamble, we have assumed that the prize is an indivisible good. However, since we assume that contestants are risk neutral, all of our results apply equally well to contests in which the prize is a divisible good and the contest success function determines the proportion of the prize received by each group. As in the case of an indivisible prize, the prize can be a public good within groups, or a private good provided we assume that members of groups have pre-committed to a rule for sharing any winnings. We can even imagine that sharing is determined in a second stage game, once the contest has been won. In that case, our analysis applies to the first stage of a subgame perfect equilibrium. For example, Katz and Tokatlidu (1996) examine the case where the sharing is determined in a second intra-group rent-seeking contest, though to simplify the analysis they assume only two groups and linear production functions.[2] The model is formally the same in all these cases, but for expositional simplicity, we refer to the indivisible, public-good case except where explicitly stated otherwise.

2 Notation and Payoffs

We study a contest played by n groups in which group i contains $N_i \geq 1$ contestants. Groups compete to win an indivisible prize which is won by one group. Each member of each group provides input into this process and the probability that group i wins is affected positively by the inputs of members of that group and negatively by members of other groups. Specifically, we suppose that the input of the kth contestant in group i is a non-negative real number x_{ik},[3] where $k = 1, \ldots, N_i$. For each group i, it is convenient to write $\mathbf{x}_i$ for the vector $(x_{i1}, \ldots, x_{iN_i})$ and $\mathbf{x}$ for the strategy profile $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. For any profile $\mathbf{x} \neq \mathbf{0}$, we assume that the probability that group i wins is given by the generalized logistic contest success function:

$$\rho_i(\mathbf{x}) = f_i(\mathbf{x}_i) / \sum_{j=1}^{n} f_j(\mathbf{x}_j),$$

where the real-valued functions $f_1, \ldots, f_n$ satisfy the following assumptions.

A1 For $i = 1, \ldots, n$, the function f_i is continuously differentiable,[4] concave, strictly increasing ($\mathbf{x}_i \geq \mathbf{x}'_i, \mathbf{x}_i \neq \mathbf{x}'_i \implies f_i(\mathbf{x}_i) > f_i(\mathbf{x}'_i)$), satisfies $f_i(\mathbf{0}) = 0$ and is unbounded above.

We can view f_i as a production function for group i, which maps the input vector $\mathbf{x}_i$ of members of the members of group i into lobbying effort $y_i = f_i(\mathbf{x}_i)$ of the group. When no contestant in any group supplies input, we suppose the prize is not awarded: $\rho_i(\mathbf{0}) = 0$ for $i = 1, \ldots, n$.

We assume that the kth contestant in group i values winning the prize at $v_{ik}(> 0)$ and incurs a cost $d_{ik}(x_{ik})$ for supplying input x_{ik}. We suppose that contestants are

[2] Choi et al. (2016) study a similar model, but with the internal and external contests running simultaneously.

[3] Our analysis extends readily, but at the expense of notational complexity, to vector inputs, possibly of different dimensions. We will omit the details for reasons of clarity.

[4] We interpret derivatives for functions not defined for negative arguments as one-sided on the boundary.

risk neutral, so the payoff of the kth contestant in group i is

$$v_{ik} f_i(\mathbf{x}_i) / \sum_{j=1}^{n} f_j(\mathbf{x}_j) - d_{ik}(x_{ik}) . \tag{1}$$

This defines a simultaneous-move game, which we denote $\mathcal{C}$. We make the following assumptions about the cost functions.

A2 For $i = 1, \ldots, n$ and $k = 1, \ldots, N_i$, the function d_{ik} is continuously differentiable, convex, strictly increasing and satisfies $d_{ik}(0) = 0$.

The value v_{ik}, which the kth member of group i assigns to winning the prize could be interpreted either as her personal evaluation of the benefit of winning when the prize is a within-group public good or her share of the prize if it is a private good. Most of our analysis applies to both cases. However, when considering comparative statics of group size it is important to distinguish between these cases. Adding players to a group in the public-good case has no effect on the payoffs of incumbent players but this is not so in the private-good case. We might typically expect v_{ik} to fall if extra players join group i and, indeed, we ascribe apparently paradoxical results on group size to the consequence of sharing the prize amongst more players rather than inefficiency.

3 Decomposition

In this and subsequent sections it will prove convenient to normalize payoffs by dividing (1) by the positive number v_{ik}. Thus, the kth contestant in group i seeks to maximize

$$\pi_{ik}(\mathbf{x}) = f_i(\mathbf{x}_i) / \sum_{j=1}^{n} f_j(\mathbf{x}_j) - c_{ik}(x_{ik}),$$

where $c_{ik}(x_{ik}) = d_{ik}(x_{ik}) / v_{ik}$, for all $i = 1, \ldots, n$ and $k = 1, \ldots, N_i$. Note that Assumption **A2** holds for c_{ik}.

To motivate our results it is helpful to look at the first-order conditions for a best response by the kth member of group i in an equilibrium profile $\widetilde{\mathbf{x}}$. These can be written as

$$\widetilde{\lambda}_i \frac{\partial f_i(\widetilde{\mathbf{x}}_i)}{\partial x_{ik}} \leq c'_{ik}(\widetilde{x}_{ik}), \tag{2}$$

with equality if $\widetilde{x}_{ik} > 0$, where

$$\widetilde{\lambda}_i = \sum_{j \neq i} f_j\left(\widetilde{\mathbf{x}}_j\right) / \left[\sum_{j=1}^{n} f_j\left(\widetilde{\mathbf{x}}_j\right)\right]^2 .$$

Interpreting $\widetilde{\lambda}_i$ as a Lagrange multiplier, these are also the first-order conditions for minimizing $\sum_\ell c_{i\ell}\left(x_{i\ell}\right)$, the total normalized cost of group i subject to $f_i\left(\mathbf{x}_i\right) \geq f_i\left(\widetilde{\mathbf{x}}_i\right)$. This shows that the equilibrium strategy profile for group i produces the lobbying effort for that group at minimum aggregate cost; production of lobbying effort is constrained efficient. It also suggests that it is helpful to define a *collective cost function* for group i for all $y \geq 0$ as

$$C_i\left(y\right) = \min_{\mathbf{x}_i \geq \mathbf{0}} \left\{ \sum_{\ell=1}^{N_i} c_{i\ell}\left(x_{i\ell}\right) : f_i\left(\mathbf{x}_i\right) \geq y \right\} . \tag{3}$$

We note, in the following lemma, proved in the Appendix, that C_i inherits many of the properties of individual cost functions.

Lemma 1 *Under Assumptions **A1** and **A2**, C_i is a convex, strictly increasing function for any i and satisfies $C_i\left(0\right) = 0$.*

We can now define a game played by the groups. The game has n players in which player (group) $i\,(= 1, \ldots, n)$ chooses $y_i \geq 0$ and receives payoff

$$\frac{y_i}{Y} - C_i\left(y_i\right),$$

where $Y = \sum_{j=1}^{n} y_j$, provided $Y > 0$ and payoff 0 if $Y = 0$. We can view this simultaneous-move game as a contest in which the ith contestant supplies effort y_i at cost $C_i\left(y_i\right)$ and the contest success function takes a simple lottery form. We refer to this game as the *reduced contest* and denote it $\mathcal{D}$.

The following theorem is our central result. The proof, in the Appendix, is an elaboration of the argument following (2).

Theorem 2 (Decomposition Theorem) *Suppose Assumptions **A1** and **A2** hold. Then, $\widetilde{\mathbf{x}}$ is a Nash equilibrium of $\mathcal{C}$ if and only if*

1. *$\left(f_1\left(\widetilde{\mathbf{x}}_1\right), \ldots, f_n\left(\widetilde{\mathbf{x}}_n\right)\right)$ is a Nash equilibrium of $\mathcal{D}$, and*
2. *$\widetilde{\mathbf{x}}_i$ achieves the minimum in the definition of $C_i\left(f_i\left(\widetilde{\mathbf{x}}_i\right)\right)$ for $i = 1, \ldots, n$.*

The constrained efficiency of production implicit in the decomposition theorem raises questions about the interpretation of comparative statics of group sizes when the prize is a private good and groups commit to sharing rules before entering the contest. In particular, the group size paradox, (larger groups being less likely to win) is often ascribed partly to free riding. However, Theorem 2 suggests that such "free-riding" is not an efficiency issue. Whilst it is certainly possible that an equilibrium entails zero input by one or more members of a group, this can occur only if it is inefficient for those members to choose a positive input level.

To study this further, suppose each group delegates the choice of strategies $\mathbf{x}_i$ for each member of the group to a manager who is tasked with looking after the interests of the group as a whole. Specifically, the manager of group i is charged with maximizing the value of winning net of the total cost incurred. This entails the manager choosing $\mathbf{x}_i$ to maximize:

$$\pi_i^* (\mathbf{x}) = f_i (\mathbf{x}_i) / \sum_{j=1}^{n} f_j \left(\mathbf{x}_j\right) - \sum_{\ell=1}^{N_i} c_{i\ell} (x_{i\ell}) .$$

Observe that, if $\widetilde{\mathbf{x}}_i$ is the manager's best response to her rivals, then $\widetilde{\mathbf{x}}_i$ must solve (3) with $y = f_i (\widetilde{\mathbf{x}}_i)$, for otherwise there would be another strategy $\mathbf{x}_i$ with a greater or equal value of $f_i (\mathbf{x}_i)$ and a smaller value of $\sum_\ell c_{i\ell} (x_{i\ell})$ and this would increase π_{ik}^* contradicting the supposition that $\widetilde{\mathbf{x}}_i$ is a best response. So the set of equilibria of the game played by the managers is the same as for the original game $\mathcal{C}$. In particular, no member of a group has an incentive to deviate unilaterally from the strategy prescribed for them by the manager.

It follows that appeals to free-riding as an explanation of the nature of equilibria need care. A recent article by Kolmar and Wagener (2013) discusses costless incentive schemes for avoiding such free-riding and conclude that groups may or may not choose to use such a scheme. (The choice is made in an initial stage of the game.) However, an alternative explanation of their results arises from noting that the incentive schemes can also be viewed as increasing the value of winning or, equivalently, decreasing costs in the second (contest) stage of the game. It is not hard to verify that, even in a simple two-player contest, the option to reduce costs in a second stage of the game, may or may not be in the interests of a contestant and, since "groups" in this case consist of a single member, this cannot be explained by appeal to free-riding.

In the next section, we apply the decomposition theorem to the existence and uniqueness of equilibria and, in the following section analyze some specific cost and production functions.

4 Existence, Uniqueness and Properties of Equilibrium

It is well-known that the reduced contest $\mathcal{D}$ has a unique Nash equilibrium (see, e.g. Cornes and Hartley 2005). It follows from the decomposition theorem that the profile of group lobbying efforts in any equilibrium is unique. The following corollary provides a formal statement.

Corollary 3 *If Assumptions* ***A1*** *and* ***A2*** *hold, then* $\mathcal{C}$ *has a Nash equilibrium. If* $\widetilde{\mathbf{x}}$ *and* $\widetilde{\mathbf{x}}'$ *are both equilibria of* $\mathcal{C}$*, then* $f_i(\widetilde{\mathbf{x}}_i) = f_i(\widetilde{\mathbf{x}}'_i)$ *for* $i = 1, \ldots, n$*.*

If, furthermore, each c_{ik} is strictly convex, then $\sum_\ell c_{i\ell}(x_{i\ell})$ is a strictly convex function of $\mathbf{x}_i$. This means that the optimization problem in (3) has a unique solution. Combined with the decomposition theorem we have the following result.[5]

Corollary 4 *Suppose Assumption* ***A1*** *and holds and* d_{ik} *is continuously differentiable, strictly convex, increasing and satisfies* $d_{ik}(0) = 0$ *for all* $i = 1, \ldots, n$ *and* $k = 1, \ldots, N_i$*. Then* $\mathcal{C}$ *has a unique equilibrium.*

5 Examples

The most commonly encountered production function in the literature is the aggregative form:

$$f_i(\mathbf{x}_i) = \sum_{\ell=1}^{N_i} x_{i\ell},$$

where it is often combined with the assumption that all cost functions are linear. It is instructive to apply the decomposition theorem to that case. Without loss of generality, we can assume $d_{ik}(x_{ik}) = x_{ik}$ for all k, so that

$$C_i(y) = \min_{\mathbf{x}_i \geq \mathbf{0}} \left\{ \sum_{\ell=1}^{N_i} \frac{1}{v_{ik}} x_{i\ell} : \sum_{\ell=1}^{N_i} x_{i\ell} \geq y \right\} = \frac{y}{\overline{v}_i},$$

where $\overline{v}_i = \max_k v_{ik}$. If there is a unique member of group i who places the highest value on winning (say the first member), the minimum is achieved at $x_{i1} = y$ and $x_{ik} = 0$ for all $k \neq 1$. If group i makes a positive effort to win the contest, only the first member supplies input.

When several players have the same maximum value of v_{ik}, any $\mathbf{x}_i \geq \mathbf{0}$ with $x_{i\ell} = 0$ for the remaining players and whose components add up to y will achieve the minimum in the definition of C_i. If group i is active in equilibrium, the equilibrium strategy profile will not be unique, although equilibrium group efforts will be. To have a unique equilibrium with several members of a group contributing positive equilibrium input, we need production or cost functions, or both, to be nonlinear and we now turn to such a case in which collective cost functions can still be calculated.

[5] A careful study of (3) shows that this conclusion remains true if for each i, in addition to Assumption **A2**, c_{ik} is strictly convex for all but one of $k = 1, \ldots, N_i$.

We will suppose that production functions are homogeneous and cost functions have constant elasticity, which is the same for all group members. This means that, for any group i, there exist $\mu_i, \omega_i > 0$, such that for any strategy profile $\mathbf{x}_i$ and $\lambda > 0$, we have (without further loss of generality)

$$f_i(\lambda \mathbf{x}_i) = \lambda^{\mu_i} f_i(\mathbf{x}_i),$$
$$d_{ik}(x_{ik}) = x_{ik}^{\omega_i}$$

for any $k = 1, \ldots, N_i$. These suppositions have implications for the form of C_i. To investigate these we will write $\widehat{\mathbf{x}}_i(y)$ for a profile that achieves the minimum in the definition of $C_i(y)$. Then

$$\sum_{\ell=1}^{N_i} c_{i\ell}(\widehat{x}_{il}(y)) \leq \sum_{\ell=1}^{N_i} c_{i\ell}(x_{i\ell})$$

for all $\mathbf{x}_i$ satisfying $f_i(\mathbf{x}_i) = y$. Multiplying this inequality by λ^{ω_i/μ_i}, we have

$$\lambda^{\omega_i/\mu_i} \sum_{\ell=1}^{N_i} c_{i\ell}(\widehat{x}_{il}(y)) \leq \lambda^{\omega_i/\mu_i} \sum_{\ell=1}^{N_i} c_{i\ell}(x_{i\ell}),$$

or

$$\sum_{\ell=1}^{N_i} c_{i\ell}\left(\lambda^{1/\mu_i}\widehat{x}_{il}(y)\right) \leq \sum_{\ell=1}^{N_i} c_{i\ell}\left(\lambda^{1/\mu_i} x_{i\ell}\right),$$

for all $\mathbf{x}_i$ satisfying $f_i(\mathbf{x}_i) = y$ and therefore also for all $\lambda^{1/\mu_i}\mathbf{x}_i$ satisfying $f_i\left(\lambda^{1/\mu_i}\mathbf{x}_i\right) = \lambda y$. It follows that $\lambda^{1/\mu_i}\widehat{\mathbf{x}}_i(y)$ achieves the minimum in the definition of $C_i(\lambda y)$. Hence,

$$C_i(\lambda y) = \sum_{\ell=1}^{N_i} c_{i\ell}(\widehat{x}_{il}(\lambda y)) = \sum_{\ell=1}^{N_i} c_{i\ell}\left(\lambda^{1/\mu_i}\widehat{x}_{il}(y)\right)$$
$$= \lambda^{\omega_i/\mu_i} \sum_{\ell=1}^{N_i} c_{i\ell}(\widehat{x}_{il}(y)) = \lambda^{\omega_i/\mu_i} C_i(y).$$

It follows that group cost functions are characterized by a single parameter $C_i(1)$ (apart from μ_i and ω_i) and they take the power form: $C_i(y) = y^{\omega_i/\mu_i} C_i(1)$. When **A1** and **A2** are satisfied, f_i is concave which implies $\mu_i \leq 1$ and c_{ik} is convex, which implies $\omega_i \geq 1$. It follows that $\omega_i/\mu_i \geq 1$, confirming that C_i is convex.

As an example, consider linear cost function $c_{ik}(x) = x$ for all i and k and a Cobb-Douglas production function:

$$f_i(\mathbf{x}) = g_i \prod_{l=1}^{N_i} x_{il}^{\alpha_{il}},$$

where $g_i, \alpha_{ik} > 0$ for all k are positive and $A_i = \sum_{l=1}^{N_i} \alpha_{il} \leq 1$ to ensure concavity. In this case, $\omega_i/\mu_i = 1/A_i$ and

$$C_i(1) = A_i g_i^{-1/A_i} / \prod_{l=1}^{N_i} (\alpha_{il} v_{il})^{\alpha_{il}/A_i}.$$

With the same cost functions and the CES production function considered by Kolmar and Rommeswinkel (2013):

$$f_i(\mathbf{x}) = g_i \left[\sum_{l=1}^{N_i} \alpha_{il} x_{il}^{\gamma_i}\right]^{1/\gamma_i},$$

where all $g_i, \alpha_{ik} > 0$ for all k and $\gamma_i \leq 1$ and $\gamma_i \neq 0$, we have $\omega_i/\mu_i = 1$ and

$$C_i(1) = g_i^{-1} / \left[\sum_{l=1}^{N_i} \alpha_{il} (\alpha_{il} v_{il})^{\rho_i}\right]^{1/\rho_i},$$

where $\rho_i = \gamma_i / (1 - \gamma_i)$.

In the CES case, ω_i/μ_i is independent of i (actually equal to 1 for all i) and this is also true for Cobb-Douglas productions functions if A_i is independent of i. In such an instance we can order the marginal values of the collective cost functions and therefore also the equilibrium probabilities of groups winning the contest. If we arrange the groups so that $C_i(1) < C_j(1)$ if $i < j$, the probability that group i wins is increasing in i. If the common value of ω_i/μ_i exceeds 1, all groups are active ($y_i > 0$). If the common value is unity (for example, with linear costs and CRS production functions), it is possible that some groups may be inactive. In this case, there will be an integer $\overline{n} \geq 2$ such that groups $i = 1, \ldots, \overline{n}$ are active and any group i for which $i > \overline{n}$ is inactive. Kolmar and Rommeswinkel write down a procedure to determine $\overline{n}$ and equilibrium group efforts and probabilities.

6 Rent Dissipation

One major theme in the literature on contests is: what proportion of the value of the prize is expended in the effort, often assumed to have no other economic worth, to win the prize? In a collective contest, intra-group inefficiency would be expected to reduce rent-dissipation through a reduction in rent-seeking activity. However, since we have shown that groups provide effort efficiently, there should be no effect on rent dissipation. It is useful to note that, since we have normalized the prize to 1, rent dissipation as a proportion of the rent takes the form

$$\rho = \sum_{i=1}^{n} \sum_{\ell=1}^{N_i} c_{i\ell}\left(\widetilde{x}_{i\ell}\right) = \sum_{j=1}^{n} C_j\left(\widetilde{y}_j\right),$$

where $\widetilde{\mathbf{x}}$ is an equilibrium strategy profile, $\widetilde{\mathbf{y}}$ is an equilibrium vector of group efforts and we refer to ρ as the *dissipation ratio*. If there are multiple equilibria, the sum is the same for all equilibria. The second sum shows that we can analyze dissipation ratios just using the reduced contest $\mathcal{D}$.

When contests are played by groups, an important issue is how group size affects the dissipation ratio. In the case where production and cost functions are linear and cost functions are the same for all members of a group, there is always an equilibrium in which only one player is active in each active group. This implies that collective costs and therefore the dissipation ratio is independent of the number of players in each group.

To study the case where costs are non-linear, suppose production functions are additive: $f_i\left(\mathbf{x}_i\right) = \sum_{\ell=1}^{N_i} x_{i\ell}$ for each group i and all contestants have the same cost functions and assign the same value to the prize: $d_{ik}\left(x\right) = x^{\alpha}$ and $v_{ik} = R$ for all i and k, where $\alpha \geq 1$. Then, from (3),

$$C_i\left(y\right) = \min_{\mathbf{x}_i \geq \mathbf{0}} \left\{ \sum_{\ell=1}^{N_i} x_{i\ell}^{\alpha}/R : \sum_{\ell=1}^{N_i} x_{i\ell} \geq y \right\}.$$

If $\alpha > 1$, the minimizer in the definition of C_i must be unique and symmetric and the constraint binding, which gives $\widehat{x}_{ik} = y/N_i$ for all k. Hence, $C_i\left(y\right) = y^{\alpha}/RN_i^{\alpha-1}$. If $\alpha = 1$, there are multiple optimal solutions but the formula for C_i remains valid.

The simplest case to consider is where $N_i = N$ for all i in which case the reduced game $\mathcal{D}$ is symmetric. It is then straightforward to see that the first-order conditions for best responses imply that the equilibrium value of Y is $\widetilde{Y}$, where

$$1 - \frac{1}{n} = \frac{\alpha \widetilde{Y}^{\alpha}}{Rn^{\alpha-1}N^{\alpha-1}}.$$

Since the equilibrium is symmetric, $\widetilde{y}_i = \widetilde{Y}/n$ for all i and therefore

$$C_i\left(\widetilde{y}_i\right) = C_i\left(\frac{\widetilde{Y}}{n}\right) = \left(1 - \frac{1}{n}\right)\frac{1}{n\alpha}.$$

It follows that the dissipation ratio is $\rho = \left(n-1\right)/n\alpha$. Intriguingly, this is independent of the number of members in each group. Furthermore, as $n \longrightarrow \infty$, the dissipation ratio approaches $1/\alpha$ (from below). This is the same dissipation ratio as in a conventional contest in which all contestants have cost function x^{α}. It is not hard to see that this limiting result remains true even if groups differ in size, provided that the number of groups of each size approaches infinity.

7 Comparative Statics

In this section, we show how the decomposition theorem can be used to study comparative statics. Collective contests are rich enough to allow for many possible studies, but here we confine ourselves to two. What are the effects of (1) adding a new group to the contest and (2) of adding members to an existing group? The general approach is to start by examining the effect of the change on collective cost functions and then investigate the effect of this change on the reduced contest. For example, adding an active group reduces the probability of an incumbent group winning the reduced contest and does not increase payoffs in equilibrium, at least under plausible assumptions on cost and production functions. To investigate the effects of such results on individual members of groups it is useful to know how an increase of group lobbying effort is reflected in the inputs of members of that group. It turns out that, under the following strengthening of Assumption **A1** and with convex costs, individual inputs rise or remain the same in equilibrium.

A1* For $i = 1, \ldots, n$,

$$f_i(\mathbf{x}_i) = \phi_i\left(\sum_{\ell=1}^{N_i} g_{i\ell}(x_{i\ell})\right), \tag{4}$$

where each ϕ_i and each g_{ik} is continuously differentiable, concave, strictly increasing, satisfies $\phi_i(0) = g_{ik}(0) = 0$ for $k = 1, \ldots, N_i$ and f_i is unbounded above.

Under this assumption, we have the following result proved in the Appendix.

Lemma 5 *Suppose Assumptions **A1*** and **A2** hold and, in addition, all cost functions in group i are strictly convex. If $y < y^*$, then the optimal solution in (3) satisfies $\widehat{\mathbf{x}}_i(y) \leq \widehat{\mathbf{x}}_i(y^*)$.*

Note that the lemma implies that, if $\widehat{x}_{ik}(y^*) = 0$, then $\widehat{x}_{ik}(y) = 0$. Active contestants do not become inactive, when the lobbying effort of the group of which they are a member increases.

7.1 Adding a Group

We first consider the effect on incumbent groups of new groups joining the contest. So, consider a collective contest in which the equilibrium value of Y in the reduced contest $\mathcal{D}$ is $\widetilde{Y}$ and suppose a new group $n + 1$ joins the contest. If at least one member of the new group is active in the new equilibrium, then the equilibrium value of Y rises to $\widetilde{Y}^*$, say. This was shown in Cornes and Hartley (2005), where it was also proved that the probabilities of incumbent groups winning the prize fall as do payoffs. The latter result can be written:

$$\frac{\widetilde{y}_i^*}{\widetilde{Y}^*} - C_i\left(\widetilde{y}_i^*\right) < \frac{\widetilde{y}_i}{\widetilde{Y}} - C_i\left(\widetilde{y}_i\right), \tag{5}$$

where $\widetilde{y}_i$ ($\widetilde{y}_i^*$) is the equilibrium effort of group $i \le n$ before (after) entry. However, without further information on cost functions, we cannot say whether $\widetilde{y}_i^*$ is bigger or smaller than $\widetilde{y}_i$.

To examine the implications for members of groups, we need to consider the two cases separately. If we have $\widetilde{y}_i^* \ge \widetilde{y}_i$, then for any member k of group i the preceding lemma gives $\widetilde{x}_{ik}^* \ge \widetilde{x}_{ik}$, where $\widetilde{x}_{ik}$ ($\widetilde{x}_{ik}^*$) is the equilibrium input of k before (after) entry. Since cost functions are increasing, this means that costs do not fall and the probability of winning does fall with entry. Hence, individual payoffs must also fall.

In the alternative case in which $\widetilde{y}_i^* < \widetilde{y}_i$, then $c_{ik}\left(\widetilde{x}_{ik}^*\right) \le c_{ik}\left(\widetilde{x}_{ik}\right)$ for all k and (5) gives

$$\begin{aligned}\sum_{\ell=1}^{N_i}\left[c_{i\ell}\left(\widetilde{x}_{i\ell}\right) - c_{i\ell}\left(\widetilde{x}_{i\ell}^*\right)\right] &= C_i\left(\widetilde{y}_i\right) - C_i\left(\widetilde{y}_i^*\right) \\ &< \frac{\widetilde{y}_i}{\widetilde{Y}} - \frac{\widetilde{y}_i^*}{\widetilde{Y}^*}.\end{aligned}$$

Since the left-hand side of this inequality is a sum of non-negative terms, we deduce that each term in this sum is less than the right-hand side. This implies

$$\frac{\widetilde{y}_i^*}{\widetilde{Y}^*} - c_{ik}\left(\widetilde{x}_{ik}^*\right) < \frac{\widetilde{y}_i}{\widetilde{Y}} - c_{ik}\left(\widetilde{x}_{ik}\right)$$

for all k. Since $\widetilde{y}_i/\widetilde{Y}$ is the probability that group i wins and the value of the prize is normalized to 1, this says that payoffs fall in this case also. That is, although the cost of the input of the kth member of group i decreases, this is more than offset by the fall in the expected value of the prize.

The following proposition summarizes these results.

Proposition 6 *Suppose additional groups join a collective contest $\mathcal{C}$ and at least one is active in the equilibrium of the enlarged contest in which Assumptions* ***A1**** *and* ***A2*** *hold. Then, inactive incumbent groups remain inactive and the equilibrium* *payoff of members of incumbent groups in which cost functions are strictly convex falls.*

7.2 Adding Contestants

It is also interesting to consider the implications of new members entering existing groups when entry does not affect valuations (the public-good case). Under the following further strengthening of **A1**, we can show that the group's marginal cost reduces.

A1** For $i = 1, \ldots, n$,

$$f_i(\mathbf{x}_i) = \sum_{\ell=1}^{N_i} g_{i\ell}(x_{i\ell}),$$

where each g_{ik} is continuously differentiable, concave, strictly increasing, satisfies $g_{ik}(0) = 0$ for $k = 1, \ldots, N_i$ and f_i is unbounded above.

Under this assumption (and **A2**), marginal group costs fall as new members join the group. The following lemma, proved in the Appendix, gives a formal statement.

Lemma 7 *Suppose C_i is defined as in (3) with $N_i \geq 2$ and Assumptions **A2** and **A1**** hold. If*

$$D_i(y) = \min_{\mathbf{x}_i} \left\{ \sum_{\ell=2}^{N_i} c_{i\ell}(x_{i\ell}) : \sum_{\ell=2}^{N_i} g_{i\ell}(x_{i\ell}) \geq y \right\},$$

then $C_i'(y) \leq D_i'(y)$.

Adding players to group i reduces the marginal group cost of group i and therefore (because $C_i(0) = 0$) actual costs as well. It follows that group i has an increased probability of winning the reduced contest $\mathcal{D}$ whilst all other groups are less likely to win and face reduced equilibrium payoffs. For further details of such comparative statics exercises, see Cornes and Hartley (2005).

We can apply these observations in the manner of the previous section to get the following result.

Proposition 8 *Suppose additional contestants join a group in a collective contest $\mathcal{C}$ and Assumptions **A2** and **A1**** hold in the enlarged contest. Then, the probability of that group winning does not fall and original members are no worse off. Furthermore, every other group faces a (weakly) diminished probability of winning and all members of other groups are no better off.*

Of course, the conclusion of this proposition depends critically on the assumption that valuations do not alter when extra members join a group. In the case where, for example, the prize is a private good which is divided equally amongst members of the winning group, the increased probability of being a member of the winning group is offset by the reduced share of the prize and which effect dominates will depend on fine details of cost and production functions. Nevertheless, the decomposition theorem offers a useful tool for studying such issues, though for reasons of space we do not do so here.

8 Conclusion

In this article, we have presented a theorem on collective contests which shows that they can be decomposed into a cost minimization problem for each group and a reduced contest between the groups. This theorem clarifies the analysis of existence, uniqueness and comparative statics and simplifies the study of rent dissipation. For example, although omitted here, we can use the theorem to explore how the internal structure of groups affects the group cost function and thereby the group's probability of winning. Furthermore, the theorem is readily extended to the case where some or all players have multi-dimensional strategy spaces, and allows us to extend strategic models such as those used to study ethnic conflict by Esteban and Ray (2008).

Finally, we note that the decomposition theorem relies heavily on risk neutrality of contestants or equivalently on having divisible prizes. Indeed, pure separation breaks down if contestants are no longer risk neutral and alternative methods must then be used. However, the constrained efficiency of intra-group strategy profiles continues to hold.

Appendix

Proof of Lemma 1 Since Assumptions **A1** and **A2** imply that we are minimizing a continuous and strictly increasing function on a closed set, the minimum is achieved, though not necessarily uniquely. We will write $\widehat{\mathbf{x}}_i(y)$ for a minimizing $\mathbf{x}_i$ in (3).

To show that C_i is strictly increasing we first observe that, since f_i is increasing (by Assumption **A1**), $f_i\left(\widehat{\mathbf{x}}_i(y)\right) = y$. Now suppose that $y^0 \in (0, y)$, which entails $f_i\left(\widehat{\mathbf{x}}_i(y)\right) > y^0$. Since f_i is continuous, there must be an $\mathbf{x}_i^0$ satisfying $\mathbf{x}_i^0 \leq \widehat{\mathbf{x}}_i(y)$ and $f_i\left(\mathbf{x}_i^0\right) \geq y^0$. Since each $c_{i\ell}$ is strictly increasing, we have

$$C_i\left(y^0\right) \leq \sum_{\ell=1}^{N_i} c_{i\ell}\left(x_{i\ell}^0\right) < \sum_{\ell=1}^{N_i} c_{i\ell}\left(\widehat{x}_{i\ell}(y)\right) = C_i(y).$$

To prove convexity, observe that, if $\mu \in (0, 1)$, concavity of f_i implies that

$$f_i\left(\mu\widehat{\mathbf{x}}_i(y) + \mu^0\widehat{\mathbf{x}}_i\left(y^0\right)\right) \geq \mu y + \mu^0 y^0,$$

where $\mu^0 = 1 - \mu$. Convexity of $c_{i\ell}$ implies that:

$$\begin{aligned} C_i\left(\mu y + \mu^0 y^0\right) &\leq \sum_{\ell=1}^{N_i} c_{i\ell}\left(\mu\widehat{x}_{i\ell}(y) + \mu^0\widehat{x}_{i\ell}\left(y^0\right)\right) \\ &\leq \sum_{\ell=1}^{N_i} \mu c_{i\ell}\left(\widehat{x}_{i\ell}(y)\right) + \mu^0 c_{i\ell}\left(\widehat{x}_{i\ell}\left(y^0\right)\right) \\ &= \mu C_i(y) + \mu^0 C_i\left(y^0\right). \end{aligned}$$

The assertion that C_i satisfies $C_i(0) = 0$ is an immediate consequence of Assumptions **A1** and **A2**. ■

In the proof of Theorem 2, it will prove convenient to use the following lemma.

Lemma 9 *Suppose Assumptions **A1** and **A2** hold and $y > 0$. If $i = 1, \ldots, n$ and $\widehat{\mathbf{x}}_i(y)$ achieves the minimum in (3) and $k = 1, \ldots, N_i$, we have*

$$C_i'(y) \leq c_{ik}'\left(\widehat{x}_{ik}(y)\right) / \frac{\partial f_i}{\partial x_{ik}}\left(\widehat{\mathbf{x}}_i(y)\right),$$

with equality if $\widehat{x}_{ik}(y) > 0$. Furthermore,

$$C_i'(0) \geq \min_{k=1,\ldots,N_i}\left\{c_{ik}'(0) / \frac{\partial f_i}{\partial x_{ik}}(\mathbf{0})\right\}.$$

Proof For any $y > 0$, the assumption that f_i is unbounded above means that there is some $\mathbf{x}_i^0$ for which $f_i\left(\mathbf{x}_i^0\right) > y$. This says that the Slater constraint qualification for the optimization problem in (3) holds (cf. Rockafellar 1972), which means that there is a Lagrange multiplier $\lambda \geq 0$ such that the (necessary) first-order conditions for this optimization problem read:

$$c_{ik}'\left(\widehat{x}_{ik}(y)\right) \geq \lambda\frac{\partial f_i}{\partial x_{ik}}\left(\widehat{\mathbf{x}}_i(y)\right),$$

with equality if $\widehat{x}_{ik} > 0$. Furthermore, marginal group cost is the slope of the minimum function and is therefore equal to the Lagrange multiplier: $C_i'(y) = \lambda$. This observation completes the proof that the first displayed inequality is necessary and sufficient.

To complete the proof, note that Assumption **A2** (specifically convexity and zero cost of zero input) applied to c_{ik} implies that $c_{ik}(x_{ik}) \geq x_{ik}c'_{ik}(0)$ for all $x_{ik} \geq 0$. Similarly, from **A1** we have

$$f_i(\mathbf{x}_i) \leq \sum_{\ell=1}^{N_i} x_{i\ell} \frac{\partial f_i}{\partial x_{i\ell}}(\mathbf{0}).$$

Hence, for any $y > 0$, we have

$$\begin{aligned} C_i(y) &\geq \min_{\mathbf{x}_i} \left\{ \sum_{\ell=1}^{N_i} x_{i\ell} c'_{i\ell}(0) : \sum_{\ell=1}^{N_i} x_{i\ell} \frac{\partial f_i}{\partial x_{i\ell}}(\mathbf{0}) \geq y \right\} \\ &= y \min_k \left\{ c'_{ik}(0) / \frac{\partial f_i}{\partial x_{ik}}(\mathbf{0}) \right\}. \end{aligned}$$

(Note that Assumption **A1** implies $(\partial f_i / \partial x_{ik})(\mathbf{0}) > 0$.) Since $C_i(0) = 0$, we can divide by y and let $y \longrightarrow 0$ to obtain the second inequality in the statement of the lemma. ■

Proof of Theorem 2 Throughout the proof, we use the fact that the convexity/concavity conditions from Assumptions **A1** and **A2** and Lemma 1 mean that first order conditions are necessary and sufficient to characterize best responses.

To prove sufficiency in the Separation Theorem, suppose $\widetilde{\mathbf{x}}$ satisfies requirements 1 and 2. Since the first of these says that $(f_1(\widetilde{\mathbf{x}}_1), \ldots, f_n(\widetilde{\mathbf{x}}_n))$ is an equilibrium of $\mathcal{D}$, the first-order conditions for best responses give

$$C'_i(\widetilde{y}_i) \geq \sum_{j \neq i} \widetilde{y}_j \left[\sum_{j=1}^{n} \widetilde{y}_j \right]^{-2},$$

with equality if $\widetilde{y}_i > 0$, where $\widetilde{y}_i = f_i(\widetilde{\mathbf{x}}_i)$ for each $i = 1, \ldots, n$. Requirement 2 says that $\widetilde{\mathbf{x}}_i$ achieves the minimum in the definition of $C_i(\widetilde{y}_i)$ and Lemma 1 implies that, for any k,

$$C'_i(\widetilde{y}_i) \leq c'_{ik}(\widetilde{x}_{ik}) / \frac{\partial f_i}{\partial x_{ik}}(\widetilde{\mathbf{x}}_i)$$

with equality if $\widetilde{x}_{ik} > 0$. Combining these inequalities gives

$$\sum_{j \neq i} f_j(\widetilde{\mathbf{x}}_j) \left[\sum_{j=1}^{n} f_j(\widetilde{\mathbf{x}}_j) \right]^{-2} \frac{\partial f_i}{\partial x_{ik}}(\widetilde{\mathbf{x}}_i) \leq c'_{ik}(\widetilde{x}_{ik}), \tag{6}$$

with equality if $\widetilde{x}_{ik} > 0$. These are the first-order conditions for best responses in $\mathcal{C}$ and show that $\mathbf{x}^*$ is a Nash equilibrium.

To prove necessity, let $\widetilde{\mathbf{x}}$ be a Nash equilibrium of $\mathcal{C}$ and write $\widetilde{y}_i = f_i(\widetilde{\mathbf{x}}_i)$ for $i = 1, \ldots, n$. If $\widetilde{y}_i > 0$, the first order conditions for best responses for members of group i in $\mathcal{C}$ are (6) and

$$\lambda = \sum_{j \neq i} f_j(\widetilde{\mathbf{x}}_j) \left[\sum_{j=1}^{n} f_j(\widetilde{\mathbf{x}}_j) \right]^{-2},$$

these can be expressed as

$$c'_{ik}(\widetilde{x}_{ik}(y)) \geq \lambda \frac{\partial f_i}{\partial x_{ik}}(\widetilde{\mathbf{x}}_i(y)),$$

with equality if $\widehat{x}_{ik} > 0$. Since $\lambda \geq 0$, these are the first order conditions for the minimization problem in (3). These conditions are necessary and sufficient by Assumptions **A1** and **A2**, which means $\widetilde{\mathbf{x}}_i$ achieves the minimum in the definition of $C_i(f_i(\widetilde{\mathbf{x}}_i))$. It follows from Lemma 1 that

$$C'_i(\widetilde{y}_i) = c'_{ik}(\widetilde{x}_{i\ell}) / \frac{\partial f_i}{\partial x_{ik}}(\widetilde{\mathbf{x}}_i).$$

Combining these observations, we get

$$C'_i(\widetilde{y}_i) = \sum_{j \neq i} \widetilde{y}_j \left[\sum_{j=1}^{n} \widetilde{y}_j \right]^{-2}.$$

These are the first order conditions for $\widetilde{y}_i$ being a best response in $\mathcal{D}$.

In the case $\widetilde{y}_i = 0$, we must have $\widetilde{\mathbf{x}}_i = \mathbf{0}$ (which achieves the minimum in (3)) and the first-order conditions imply that, for all k,

$$\sum_{j \neq i} f_j(\widetilde{\mathbf{x}}_j) \left[\sum_{j=1}^{n} f_j(\widetilde{\mathbf{x}}_j) \right]^{-2} \frac{\partial f_i}{\partial x_{ik}}(\mathbf{0}) \leq c'_{ik}(0).$$

Combined with the second inequality in Lemma 1, we deduce

$$\sum_{j \neq i} \widetilde{y}_j \left[\sum_{j=1}^{n} \widetilde{y}_j \right]^{-2} \leq C'_i(0),$$

which implies that 0 is a best response by group i in $\mathcal{D}$. We have demonstrated that $\widetilde{\mathbf{y}}$ is an equilibrium of $\mathcal{D}$. ■

Proof of Lemma 5 To see that $c_{ik}\left(\widehat{x}_{ik}\left(y\right)\right)$ is increasing in y when Assumption **A1*** holds, we start by noting that the group cost function can be rewritten:

$$C_i\left(y\right) = \min_{\mathbf{x}_i}\left\{\sum_{\ell=1}^{N_i} c_{i\ell}\left(x_{i\ell}\right) : \sum_{\ell=1}^{N_i} g_{i\ell}\left(x_{i\ell}\right) \geq \phi_i^{-1}\left(y\right)\right\}, \tag{7}$$

where ϕ_i^{-1} is the inverse function of ϕ_i. Convexity of the cost functions and concavity of the $g_{i\ell}$ imply the existence of a Lagrange multiplier $\lambda_i\left(y\right) \geq 0$ such that

$$\sum_{\ell=1}^{N_i} c_{i\ell}\left(\widehat{x}_{i\ell}\left(y\right)\right) - \lambda_i\left(y\right)\sum_{\ell=1}^{N_i} g_{i\ell}\left(\widehat{x}_{i\ell}\left(y\right)\right) < \sum_{\ell=1}^{N_i} c_{i\ell}\left(x_{i\ell}\right) - \lambda_i\left(y\right)\sum_{\ell=1}^{N_i} g_{i\ell}\left(x_{i\ell}\right) \tag{8}$$

for any $\mathbf{x}_i \geq \mathbf{0}$ satisfying $\mathbf{x}_i \neq \widehat{\mathbf{x}}_i\left(y\right)$ the constraint in (7). Now suppose $y^* > y$.

It is possible that $\widehat{\mathbf{x}}_i\left(y\right) = \widehat{\mathbf{x}}_i\left(y^*\right)$. If not, combining (8) with the constraint in (7), we have

$$\sum_{\ell=1}^{N_i} c_{i\ell}\left(\widehat{x}_{i\ell}\left(y\right)\right) - \lambda_i\left(y\right)\phi_i^{-1}\left(y\right) < \sum_{\ell=1}^{N_i} c_{i\ell}\left(\widehat{x}_{i\ell}\left(y^*\right)\right) - \lambda_i\left(y\right)\phi_i^{-1}\left(y^*\right),$$

$$\sum_{\ell=1}^{N_i} c_{i\ell}\left(\widehat{x}_{i\ell}\left(y^*\right)\right) - \lambda_i\left(y^*\right)\phi_i^{-1}\left(y^*\right) < \sum_{\ell=1}^{N_i} c_{i\ell}\left(\widehat{x}_{i\ell}\left(y\right)\right) - \lambda_i\left(y^*\right)\phi_i^{-1}\left(y\right),$$

where we have used the fact that $\sum_l g_{il}\left(\widehat{x}_{il}\left(y\right)\right) = \phi_i^{-1}\left(y\right)$ (since g_i is increasing). Adding these inequalities shows that

$$\left[\phi_i^{-1}\left(y^*\right) - \phi_i^{-1}\left(y\right)\right]\left[\lambda_i\left(y^*\right) - \lambda_i\left(y\right)\right] > 0.$$

Since ϕ_i is strictly increasing, we have $\phi_i^{-1}\left(y^*\right) > \phi_i^{-1}\left(y\right)$ and may conclude that $\lambda_i\left(y^*\right) > \lambda_i\left(y\right)$.

It follows from (4) and (8) that $\widehat{x}_{ik}\left(y\right)$ minimizes $c_{ik}\left(x_{i\ell}\right) - \lambda_i\left(y\right)g_{ik}\left(x_{ik}\right)$ subject to $x_{ik} \geq 0$ and therefore, for any k for which $\widehat{x}_{ik}\left(y\right) \neq \widehat{x}_{ik}\left(y^*\right)$, we have

$$c_{ik}\left(\widehat{x}_{ik}\left(y\right)\right) - \lambda_i\left(y\right)g_{ik}\left(\widehat{x}_{ik}\left(y\right)\right) < c_{ik}\left(\widehat{x}_{ik}\left(y^*\right)\right) - \lambda_i\left(y\right)g_{ik}\left(\widehat{x}_{ik}\left(y^*\right)\right),$$

$$c_{ik}\left(\widehat{x}_{ik}\left(y^*\right)\right) - \lambda_i\left(y^*\right)g_{ik}\left(\widehat{x}_{ik}\left(y^*\right)\right) < c_{ik}\left(\widehat{x}_{ik}\left(y\right)\right) - \lambda_i\left(y^*\right)g_{ik}\left(\widehat{x}_{ik}\left(y\right)\right).$$

In the case $\lambda_i(y) > 0$, dividing the first inequality by $\lambda_i(y)$, the second by $\lambda_i(y^*)$, adding the results and rearranging gives

$$\left[\frac{1}{\lambda_i(y)} - \frac{1}{\lambda_i(y^*)}\right]\left[c_{ik}\left(\widehat{x}_{ik}\left(y^*\right)\right) - c_{ik}\left(\widehat{x}_{ik}(y)\right)\right] > 0.$$

Since we have already shown that $\lambda_i(y^*) > \lambda_i(y)$, we can deduce that $c_{ik}(\widehat{x}_{ik}(y^*)) > c_{ik}(\widehat{x}_{ik}(y))$. Since c_{ik} is a strictly increasing function, this implies $\widehat{x}_{ik}(y^*) > \widehat{x}_{ik}(y)$. In the case $\lambda(y) = 0$, we have $\widehat{x}_{ik}(y) = 0$ and $c_{ik}(\widehat{x}_{ik}(y^*)) > 0 = c_{ik}(\widehat{x}_{ik}(y))$. Hence, $\widehat{x}_{ik}(y^*) > \widehat{x}_{ik}(y)$. ■

Proof of Lemma 7 By definition,

$$C_i(y) = c_{i1}(\widehat{x}_{i1}(y)) + D_i(y - g_{i1}(\widehat{x}_{i1}(y)))$$

for any $y \geq 0$, where $\widehat{x}_{ik}(y)$ is the optimal solution of (3). For any $y' > y$, the definition of C_i implies

$$\begin{aligned} C_i\left(y'\right) &= c_{i1}\left(\widehat{x}_{i1}\left(y'\right)\right) + D_i\left(y' - g_{i1}\left(\widehat{x}_{i1}\left(y'\right)\right)\right) \\ &\leq c_{i1}(\widehat{x}_{i1}(y)) + D_i\left(y' - g_{i1}(\widehat{x}_{i1}(y))\right). \end{aligned}$$

Hence,

$$\begin{aligned} C_i\left(y'\right) - C_i(y) &\leq D_i\left(y' - g_{i1}(\widehat{x}_{i1}(y))\right) - D_i(y - g_{i1}(\widehat{x}_{i1}(y))) \\ &\leq D_i\left(y'\right) - D_i(y), \end{aligned}$$

using convexity of D_i (Lemma 1) and the fact that $g_{i1}(\widehat{x}_{i1}(y)) > 0$. Dividing by $y' - y$ and letting $y' \longrightarrow y$ gives the result. ■

References

Baik, K. H. (1993). Effort levels in contests: The public-good prize case. *Social Choice and Welfare, 30*, 103–117.

Baik, K. H. (2008). Contests with group-specific public-good prizes. *Social Choice and Welfare, 30*, 103–117.

Baik, K. H., Kim, I.-G., & Na, S. (2001). Bidding for a group-specific public good prize. *Journal of Public Economics, 82*, 415–429.

Choi, J. P., Chowdhury, S. M., & Kim, J. (2016). Group contests with internal conflict and power asymmetry. *Scandinavian Journal of Economics 82*, 816–840.

Chowdhury, S. M., Lee, D., & Sheremeta, R. M. (2013). Top guns may not fire. Best-shot group contests with group-specific public good prizes. *Journal of Economic Behavior and Organization, 92*, 94–103.

Cornes, R. C., & Hartley, R. (2005). Asymmetric contests with general technologies. *Economic Theory, 26*, 923–946.

Cornes, R. C., & Hartley, R. (2007). Weak links, good shots and other public good games: Building on BBV. *Journal of Public Economics, 91*, 1684–1707.
Dijkstra, B. R. (1998). Cooperation by way of support in a rent seeking contest for a public good. *European Journal of Political Economy, 14*, 703–725.
Epstein, G. S., & Mealam, Y. (2009). Group specific public goods, orchestration of interest groups with free riding. *Public Choice, 139*, 357–369.
Esteban, J., & Ray, D. (2001). Collective action and the group size paradox. *American Political Science Review, 95*, 663–672.
Esteban, J., & Ray, D. (2008). On the salience of ethnic conflict. *American Economic Review, 98*, 2185–2202.
Katz, E., Nitzan S., & Rosenberg, J. (1990). Rent-seeking for pure public goods. *Public Choice, 65*, 49–60.
Katz, E., & Tokatlidu, J. (1996). Group competition for rents. *European Journal of Political Economy, 12*, 599–607.
Kolmar, M., & Rommeswinkel, H. (2013). Contests with group-specific public goods and complementarities in efforts. *Journal of Economic Behavior and Organization, 89*, 9–22.
Kolmar, M., & Wagener, A. (2013). Inefficiency as a strategic device in group contests against dominant opponents. *Economic Inquiry, 51*, 2083–2095.
Nitzan, S. (1991). Collective rent dissipation. *Economic Journal, 101*, 1522–1534.
Nitzan, S., & Ueda, K. (2011). Prize sharing and collective contests. *European Economic Review, 55*, 678–687.
Nitzan S., & Ueda, K. (2014). Intra-group heterogeneity in collective contests. *Social Choice and Welfare, 43*, 219–238.
Olsen, M. (1965). *The logic of collective action*. Cambridge: Harvard University Press.
Riaz, K., Shogren, J. F., & Johnson, S. R. (1995). A general model of rent-seeking for public goods. *Public Choice, 82*, 243–259.
Rockafellar, R. T. (1972). *Convex analysis*. Princeton: Princeton University Press.
Ryvkin, D. (2011). The optimal sorting of players in contests between groups. *Games and Economic Behavior, 73*, 564–472.
Topolyan, I. (2014). Rent-seeking for a public good with additive contributions. *Social Choice and Welfare, 42*, 465–476.
Ueda, K. (2002). Oligopolization in collective rent-seeking. *Social Choice and Welfare, 19*, 613–626.

Multiple-Aggregate Games

Alex Dickson

1 Introduction

Richard and I had many interesting discussions about aggregative games and games in which there are multiple aggregates, and it was firmly on both of our agendas to pursue joint research on multiple aggregate games. As is often the case, momentum in pursuing ideas takes time to establish, and it was very unfortunate that we were not afforded the opportunity to work closely on developing these ideas before Richard sadly passed away in 2015. It is a great honour to have the opportunity to communicate the current state of my thoughts on multiple aggregate games in this volume dedicated to Richard: the aim has been to provide an accessible exposition of the ideas and establish a framework for analysis, rather than to derive the most general results under the weakest assumptions, that I believe Richard would appreciate. The work has undoubtedly benefitted from the discussions I had with Richard, as well as with Roger Hartley, and it is also without doubt that what is presented here is inferior to what might have been achieved had Richard co-authored the contribution: he had amazing intuition and an ability to explain complex ideas in a simple way, that only comes from having a truly deep understanding. I hope to have done the ideas justice.

In a strategic decision making environment there is strategic interdependence between individuals playing a non-cooperative game; each is influenced by, and influences, the other players in the game. Sometimes individuals care about exactly which of their adversaries does what, but in many interesting economic applications players care only about the aggregation of other players' actions, since it is this that influences their payoff. Such games are called *aggregative games*. Often,

---

A. Dickson (✉)
Department of Economics, Strathclyde Business School, University of Strathclyde, 199 Cathedral Street, Glasgow G4 0QU, UK
e-mail: alex.dickson@strath.ac.uk

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_3

individuals within an environment are organised into groups and they contribute to the collective action of their group which in part determines their payoff, but they are also affected by the collective actions of other groups: there are externalities between groups that are transmitted through the aggregation of groups' actions. Whilst the theory of aggregative games has been successfully applied to study games with a single aggregate, the setting just described features multiple aggregations of actions, one for each group, and the nature of the intra-group strategic interaction—where players contribute to the collective action of their group—may be very different to the inter-group strategic interaction. The aim of this paper is to establish a framework in which to consider such 'multiple aggregate games'; present a method to analyse the existence and properties of equilibria; and to discuss some applications of the theory—to contests; public goods games; and bilateral oligopoly—to demonstrate how useful the technique is for analysing strategic interactions between groups of individuals.

Consider a simultaneous-move game of complete information involving $i = 1, \ldots, N$ individuals where each has to decide on a single action $x_i \in \mathbb{R}_+$. Each player's payoff depends on their own action and the vector of all other players' actions. The game is an 'aggregative game' if each player cares only about the aggregation of other players' actions, $X_{-i} = X - x_i$, where X is the sum of all players' actions. The standard approach to finding a Nash equilibrium involves identifying each player's best response function, $b_i(X_{-i})$; this is player i's action consistent with a Nash equilibrium in which the aggregate actions of other players is X_{-i}. The problem is that for each player this is defined on a different domain, and therefore the joint best response function, of which a fixed point must be found, has as many dimensions as there are players. An aggregative approach does something different: rather than finding a best response, instead consider the action of player i consistent with a Nash equilibrium in which the aggregate of *all* players, including player i, takes the value X. This gives the 'replacement function'[1] $r_i(X)$, so called because finding it involves replacing X_{-i} with $X - x_i$ in the equation that defines the best response and then solving for x_i. A Nash equilibrium is identified at the level of the aggregation of actions, and requires aggregate consistency between individual actions and the aggregate; that is, for X to be such that the sum of replacement functions evaluated at X exactly equals X. This is a much simpler problem than finding mutually consistent best responses! Existence, uniqueness and comparative static properties can be investigated by understanding the properties of replacement functions and their aggregation, which is tractable even in a game with heterogeneous players; and whether these players are active in equilibrium or not can be deduced by evaluating their replacement function at the equilibrium aggregate.

The methods of aggregative games have been used to good effect in the study of Cournot oligopoly (Novshek 1985); public goods games (Cornes and Hartley 2007);

[1] Elsewhere this is called the 'cumulative best reply' (Selten 1970), the 'inclusive reaction function' (Wolfstetter 1999) and the 'backward reaction correspondence' (Novshek 1985).

and Tullock contests (Cornes et al. 2005); among others. In contests, a contestant's *share* of the aggregate action naturally features in their payoff function, and in the analysis of the game it is often convenient to use share functions rather than replacement functions; since $s_i(X) = r_i(X)/X$ can be monotonically decreasing in X when $r_i(X)$ is not. Aggregate consistency with share functions requires their sum to take the value 1. The share function approach was first introduced by Cornes and Hartley (2000) in studying joint production games. In addition to exploring the use of aggregative games in particular applications, there have been some general treatments including those by Corchon (1994), Jensen (2010) (who considers that the aggregation of players' actions can be more general than the simple unweighted sum), Cornes and Hartley (2012), and Acemoglu and Jensen (2013).

In a 'multiple aggregate game' each player is a member of a single group and their action contributes to the collective action of their group. Individuals within a group care about their own action, the collective action of their group, and also the collective actions of other groups: there is intra-group strategic interaction which takes the form of an aggregative game; and inter-group spill-overs that transmit through the aggregate actions of groups. The applicability of this framework that extends the scope of aggregative games is clear. Inspiration for the study of multiple aggregate games comes from the analysis of bilateral oligopoly (see, for example, Dickson and Hartley 2008) in which there is a set of buyers and a set of sellers and, in essence, each group of traders plays a simple Tullock contest in which they receive a proportional share of a prize, the size of which is determined by the aggregate actions of the other group of traders.

The method used to analyse multiple aggregate games first resolves the intra-group strategic interaction, and then essentially treats groups as players that choose an aggregate action to resolve the inter-group interaction. First, a group is selected and the aggregate actions of other groups are fixed at arbitrary levels. This defines a 'partial game' that involves only the members of the selected group, and since each group member cares only about their own action and the aggregation of other group members' actions, this is an aggregative game. Within the partial game, aggregative methods can be applied to identify a Nash equilibrium: individual replacement or share functions are derived that represent the consistent behaviour of group members; then aggregate consistency within the group is imposed to identify the Nash equilibrium in the partial game, which reveals the 'group best response'. This is repeated for each group, and the resulting group best responses represent the collective action of each group consistent with a Nash equilibrium in which the aggregate actions of other groups take a particular value, having accounted for the strategic tension within groups. A Nash equilibrium in the full game can then be identified at the level of group aggregates, that requires the aggregate action of each group to be a group best response to the aggregate actions of the other groups.

Whilst identifying Nash equilibrium is a fixed point problem, it involves the joint group best response and therefore only has as many dimensions as there are groups even though there are heterogeneous players within each group. This may be as few as two, as in bilateral oligopoly. As such, exploiting the aggregative nature of the game considerably simplifies the analysis and, since group best responses are found using the aggregation of replacement or share functions whose properties are easily deduced, the features of Nash equilibrium can be easily understood.

In some games there is more structure to the group interaction: a 'nested aggregative game' has the feature that individuals not only care just about the aggregation of others' actions within their group, they also care only about the aggregation of other groups' actions. Such games are aggregative both at the level of individuals within groups, and at the level of groups. The analysis slightly differs to exploit this additional aggregative structure: within 'partial games' individual and group consistency with a Nash equilibrium in which the aggregate of all groups' actions takes a particular value is sought to define 'group replacement functions', following which overall consistency is required for a Nash equilibrium, which needs the sum of group replacement functions to be equal to the aggregate action, a simpler problem than finding mutually consistent group best responses. Thus, consistency in aggregation is required twice: once at the level of individuals within groups in partial games; and once at the level of groups within the full game. The analysis of a game between individuals using aggregative techniques renders the study of equilibrium tractable and permits uniqueness of equilibrium to be considered even in the presence of heterogeneous players, and the same is true of a nested aggregative game with heterogeneous groups and heterogeneity of players within groups, which is explored here in a general setting.

There is some existing literature on strategic interactions between individuals within groups that is related to the ideas presented here. Cornes et al. (2005) consider a model in which individuals in groups contribute to a public good enjoyed by their group, and there are also spill-overs in the public good provided by each group. Restrictions are imposed on the nature of the spill-overs that ensure the game has the form of what has been called here a nested aggregative game, and the idea of both group and overall consistency to identify Nash equilibria is introduced. Nitzan and Ueda (2014) study a 'collective contest' in which individuals in groups contribute effort to the group in contesting a rent. The cost of effort is heterogeneous among group members, as is their valuation of the rent, and the approach taken to analyse a Nash equilibrium recognises that it is a nested aggregative game and appeals to group and overall consistency to analyse the effect of heterogeneity within groups.

In strategic interactions where group structure is important, some contributions have used ideas that are similar to the partial game approach taken here. In particular, Baik (2008) studies a collective contest and uses the idea of a 'group-specific equilibrium' to analyse the game, albeit in a simple setting since there is essentially a single active player in each group. Kolmar and Rommeswinkel (2013) study a contest between groups in which there are complementarities in effort within groups, and use the idea of a group best response function to identify Nash equilibria.

In the setting explored here each individual belongs to a single group and contributes only to the aggregate of that group; there are multiple aggregates because there are multiple groups. In other settings there might be multiple aggregations of actions where all players contribute to all aggregates. Models of production and appropriation fall into this category and, whilst aggregative methods can be used to analyse such models (Cornes et al. 2010), the lack of group structure means they fall outside the remit of the current exposition.

The remainder of the paper is structured as follows. Section 2 presents the economic environment, defines the game that is played, and introduces the idea of partial games and group best responses. In Sect. 3 partial games are analysed by exploiting their aggregative properties to derive group best responses, and in Sect. 4 Nash equilibrium in the full game is studied by considering mutual consistency of group best responses. Section 5 considers the special case of nested aggregative games. In Sect. 6 some applications of the method—to contests; bilateral oligopoly; and public goods—are considered, and conclusions follow. All proofs are contained in an Appendix.

2 The Economic Framework

Consider a strategic interaction between individuals that are exogenously organised in groups where each individual belongs to a single group and their payoffs depend on the actions of their fellow group members and on the actions of members of other groups, perhaps in a fundamentally different way than within the group. It is natural to think of the influence of members of other groups working through the aggregation (i.e., the sum) of those groups' actions, and attention is restricted to this case. The (finite) set of groups is $J = \{1, \ldots, j, \ldots, N\}$ and the (finite) set of individuals in group j is $I^j = \{1, \ldots, i, \ldots, N^j\}$. Subscripts are used to identify individuals, superscripts to identify the group they belong to. Each individual must simultaneously choose a single action $x_i^j \in \mathbb{R}_+$; capitals are used to represent aggregations of actions, and vectors of actions are in boldface, as the following statement makes clear.

Notation $\mathbf{x} = \{x_i^j\}_{i \in I^j, j \in J}$ *is the vector of all players' actions.* $\mathbf{x}^j = \{x_i^j\}_{i \in I^j}$ *is the vector of all actions of members of group j, and* $\mathbf{x}_{-i}^j = \mathbf{x}^j \setminus x_i^j$. $X^j = \sum_{i \in I^j} x_i^j$ *is the aggregation of actions of the members of group j, and* $X_{-i}^j = X^j - x_i^j$. $\mathbf{X} = \{X^j\}_{j \in J}$ *is the vector of all group aggregates, and* $\mathbf{X}^{-j} = \mathbf{X} \setminus X^j$. *Where appropriate,* $X = \sum_{j \in J} X^j$ *is the aggregation of all groups' aggregate actions, and* $X^{-j} = X - X^j$.

In a strategic environment individuals' actions have external consequences for others, so typically an individual's payoff will depend both on their own action and on the actions chosen by all other individuals. Here, an individual's actions have external consequences both within and outside their group, and the effect on members of other groups comes only through the aggregation of the group's actions. As such, each individual cares about the actions of members of other groups only through their aggregation, so the payoff to a typical individual in group j can be written

$$u_i^j(x_i^j, \mathbf{x}_{-i}^j; \mathbf{X}^{-j}).$$

This allows externalities between players within a group as well as externalities between groups that work the level of aggregate group actions to be considered, where the nature of the strategic interaction within groups may be very different to that between groups: for example, an individual's actions may have negative consequences for the members of their own group, and positive consequences for the members of other groups, or vice versa.

The game $\mathcal{G}$ is the simultaneous-move game of complete information with player set $\cup_{j \in J} I^j$; actions $x_i^j \in \mathbb{R}_+$; and payoffs $u_i^j(x_i^j, \mathbf{x}_{-i}^j; \mathbf{X}^{-j})$: the equilibrium concept is Nash equilibrium in pure strategies. If no further attention was paid to the group structure of the game, the analysis would proceed by attempting to find a vector of actions, one for each player, that constitute mutually consistent best responses. This involves finding a fixed point of the joint best response function, that has as many dimensions as there are players in the game; whilst there are well-known approaches for studying the existence of such fixed points, understanding the properties of equilibrium is a more difficult task. This is particularly true when the game does not exhibit strategic complementarities so the tools of supermodular games cannot be exploited, which is likely to be the case in many interesting applications given the potentially different nature of the intra- and inter-group interaction.

By recognising the group structure of the game, a somewhat different approach to identifying Nash equilibria can be taken. The method follows a two-step procedure. First, select a group and fix the actions of the members of all other $N - 1$ groups, so the aggregate action of each of these groups (which is what the members of the group in question care about) takes a particular value. Consider the strategic interaction among the members of the group in question, seeking to find actions that are consistent with a Nash equilibrium in the game in which the other groups' aggregate actions take the specified values. Repeat this for each group (fixing the aggregate actions of the other $N - 1$ groups in turn), which identifies consistent behaviour *within* groups taking as fixed the aggregate actions of all other groups. The second step then looks at *between*-group consistency. The aggregation of the consistent individual actions found in the first step gives a 'group best response' to the aggregate actions of other groups; a Nash equilibrium of the game requires mutual consistency of these group best responses at the level of group aggregates. Once an equilibrium has been identified, individual actions can be deduced from the characterisation of equilibrium behaviour within groups from step 1, evaluated at the equilibrium values of the aggregate actions of other groups.

More precisely, select a group j and fix the aggregate actions of other groups at some levels collected in $\mathbf{X}^{-j}$. Define a 'partial game' $\mathcal{G}^j(\mathbf{X}^{-j})$ amongst the members of group j, in which the aggregate actions of all other groups are fixed. The analysis of the intra-group strategic interaction involves finding a Nash equilibrium of this partial game. Restrictions will be imposed ensuring uniqueness of equilibrium within each partial game for any $\mathbf{X}^{-j}$; the action of individual i in this Nash equilibrium is written $\tilde{x}_i^j(\mathbf{X}^{-j})$ and the aggregation of group j's equilibrium actions is $\tilde{X}^j(\mathbf{X}^{-j}) = \sum_{i \in I^j} \tilde{x}_i^j(\mathbf{X}^{-j})$. Having resolved the strategic interaction within the group, this function gives the 'group-j best response' to the aggregate actions of the other groups contained in $\mathbf{X}^{-j}$. This is then repeated for all other $N-1$ groups to deduce a group best response function for each group. The remaining task is to ensure between-group consistency, which requires mutual consistency of group best responses: to identify a Nash equilibrium in the game $\mathcal{G}$ a vector of group aggregates $\mathbf{X}^*$ is sought such that $X^{j*} = \tilde{X}^j(\mathbf{X}^{-j*})$ for all $j \in J$; the equilibrium action of each player will consequently be $x_i^{j*} = \tilde{x}_i^j(\mathbf{X}^{-j*})$. Essentially, having ensured intra-group consistency groups are treated as players in an N-player game where they 'choose' aggregate actions and have best response functions given by $\tilde{X}^j(\mathbf{X}^{-j})$.

The following proposition establishes that the two-step procedure just outlined is valid in identifying Nash equilibria in the game $\mathcal{G}$, since mutually consistent group best responses are in one-to-one correspondence with Nash equilibria in the game.

Proposition 1 *Consider the N partial games $\{\mathcal{G}^j(\mathbf{X}^{-j})\}_{j\in J}$ of the game $\mathcal{G}$ in which the player set is the members of group j, their actions are $x_i^j \in \mathbb{R}_+$, and their payoffs are $u_i^j(x_i^j, \mathbf{x}_{-i}^j; \mathbf{X}^{-j})$ where $\mathbf{X}^{-j}$ is considered fixed. Suppose a Nash equilibrium in $\mathcal{G}^j(\mathbf{X}^{-j})$ exists and is unique for any $\mathbf{X}^{-j} \in \mathbb{R}_+^{N-1}$, and write $\tilde{x}_i^j(\mathbf{X}^{-j})$ for the equilibrium strategy of player i and $\tilde{X}^j(\mathbf{X}^{-j}) = \sum_{i\in I^j} \tilde{x}_i^j(\mathbf{X}^{-j})$ for the aggregation of group j's actions in the Nash equilibrium. Then $\mathbf{x}^*$ is a Nash equilibrium in $\mathcal{G}$ if and only if $X^{j*} = \tilde{X}^j(\mathbf{X}^{-j*})$ for all $j \in J$, where $x_i^{j*} = \tilde{x}_i^j(\mathbf{X}^{-j*})$.*

This proposition supposes there is a unique Nash equilibrium in each partial game. To understand the conditions under which this will be true attention is restricted to strategic interactions where, within each group, individuals only care about the aggregation of other group members' actions. In this case, a player's payoff can (with a slight abuse of notation) be written

$$u_i^j(x_i^j, X_{-i}^j; \mathbf{X}^{-j}).$$

With this structure, which is a common feature of many games with continuous strategies, a typical individual's payoff depends on their own action, the aggregation of their group's actions, and the vector of all other groups' aggregate actions, since it can be written

$$\tilde{u}_i^j(x_i^j, X^j; \mathbf{X}^{-j}) \equiv u_i^j(x_i^j, X^j - x_i^j; \mathbf{X}^{-j}). \tag{1}$$

As such, once the vector of other groups' aggregate actions is fixed, each partial game $\mathcal{G}^j(\mathbf{X}^{-j})$ is an aggregative game, and this aggregative structure will be exploited to establish uniqueness of equilibrium within partial games. The game $\mathcal{G}$, being constituted of N aggregative games, is thus a 'multiple-aggregate game'.

In the special case of a 'nested aggregative game' individuals in group j not only care just about the aggregation of other group members' actions, they also care only about the aggregation of other groups' actions. Then payoffs (again with a slight abuse of notation) can be written

$$\tilde{u}_i^j(x_i^j, X^j; X^{-j}).$$

Since $X^{-j} = X - X^j$, in such games an individual's payoff will depend on their own action, the aggregation of their group's action, and the aggregation of all groups' actions, since

$$\hat{u}_i^j(x_i^j, X^j; X) \equiv \tilde{u}_i^j(x_i^j, X^j; X - X^j).$$

This special case is the focus of attention in Sect. 5.

3 Study of Group Partial Games

In the partial game $\mathcal{G}^j(\mathbf{X}^{-j})$ with $\mathbf{X}^{-j}$ fixed, a Nash equilibrium among the N^j members of group j is sought. The payoff functions of these players, as noted in (1), can be written $\tilde{u}_i^j(x_i^j, X^j; \mathbf{X}^{-j})$; consequently, the group-j partial game is an aggregative game since with $\mathbf{X}^{-j}$ fixed each player's payoff depends only on their own action and the aggregation of all players' actions (within that player's group) which will be exploited to study the Nash equilibria of the partial game and understand the features of group best response functions $\tilde{X}^j(\mathbf{X}^{-j})$.

Basically, the aim is to define for each player a 'share function' that represents their behaviour consistent with a Nash equilibrium in a partial game in which the group aggregate takes a particular value, and show that there is a Nash equilibrium in the partial game if and only if the sum of shares equals one. If share functions are strictly decreasing in the group aggregate then the sum of shares will inherit this property, so if the sum of shares exceeds one when the group aggregate is small, and is less than one when it is large, there will be a unique Nash equilibrium in the partial game. Under what conditions is this true?

If a player's payoff function u_i^j is strictly concave in own strategy, which will be assumed, their best response will be unique and the best response function, denoted $b_i^j(X_{-i}^j; \mathbf{X}^{-j})$, is identified by the necessary and sufficient first-order condition. Thus, $b_i^j(X_{-i}^j; \mathbf{X}^{-j}) = \max\{0, x_i^j\}$ where x_i^j satisfies

$$\frac{\partial u_i^j(x_i^j, X_{-i}^j; \mathbf{X}^{-j})}{\partial x_i^j} = 0. \tag{2}$$

A player's best response gives their action consistent with a Nash equilibrium in $\mathcal{G}^j(\mathbf{X}^{-j})$ in which the actions of all other players in group j sum to X_{-i}^j.

Rather than working with best responses, an aggregative approach considers the strategy of a player that is consistent with a Nash equilibrium in which the aggregate action of *all* members of the group, including the player in question, takes a particular value X^j. This will be given by the player's '*replacement* function' $r_i^j(X^j; \mathbf{X}^{-j})$ which is defined by

$$r_i^j(X^j; \mathbf{X}^{-j}) \equiv b_i^j(X^j - r_i^j(X^j; \mathbf{X}^{-j}); \mathbf{X}^{-j}).$$

Note that the replacement function is given by $r_i^j(X^j; \mathbf{X}^{-j}) = \max\{0, x_i^j\}$ where x_i^j is defined by

$$\frac{\partial u_i^j(x_i^j, X^j - x_i^j; \mathbf{X}^{-j})}{\partial x_i^j} = 0 \tag{3}$$

so long as $r_i^j(X^j; \mathbf{X}^{-j}) \leq X^j$, otherwise the function is deemed undefined.

A player's replacement function gives their individual action consistent with a Nash equilibrium in the game $\mathcal{G}^j(\mathbf{X}^{-j})$ in which the aggregate action of all members of group j is X^j. Consistency of individual group members' actions with the aggregate of the group requires the sum of their actions consistent with a particular group aggregate to be exactly equal to that group aggregate. This simple equilibrium identification condition in the partial game, which involves finding a fixed point of $\sum_{i \in I^j} r_i^j(X^j; \mathbf{X}^{-j}) : \mathbb{R}_+ \to \mathbb{R}_+$ at the level of the aggregate action of the group, makes clear the appeal of an aggregative game approach.

Proposition 2 $\mathbf{x}^{j*}$ *is a Nash equilibrium in the game* $\mathcal{G}^j(\mathbf{X}^{-j})$ *if and only if*

$$\sum_{i \in I^j} r_i^j(X^{j*}; \mathbf{X}^{-j}) = X^{j*}. \tag{4}$$

An aggregate action by the members of group j that satisfies the consistency condition (4) constitutes a 'group best response' in $\mathcal{G}^j(\mathbf{X}^{-j})$, which is denoted $\tilde{X}^j(\mathbf{X}^{-j})$. Whether Nash equilibria in partial games are unique, so $\tilde{X}^j(\mathbf{X}^{-j})$ can be considered a function, will rely on the monotonicity properties of the representation of consistent individual behaviour with respect to the group aggregate. Rather than working with *levels* of a player's action, represented by the replacement function, it is often more convenient to work with their *share* of the group aggregate, $\sigma_i^j = x_i^j / X^j$, as this can be monotonically decreasing in X^j when replacement functions are not. For $X^j > 0$, an individual's 'share function' is[2] $s_i^j(X^j; \mathbf{X}^{-j}) = r_i^j(X^j; \mathbf{X}^{-j})/X^j$ which is implicitly defined by $s_i^j(X^j; \mathbf{X}^{-j}) = \max\{0, \sigma_i^j\}$ where σ_i^j satisfies

$$l_i^j(\sigma_i^j, X^j; \mathbf{X}^{-j}) \equiv \frac{\partial u_i^j(\sigma_i^j X^j, X^j[1 - \sigma_i^j]; \mathbf{X}^{-j})}{\partial x_i^j} = 0 \tag{5}$$

so long as $\sigma_i^j \leq 1$, otherwise the share function is undefined.

The analogue of individual actions summing to the aggregate action that achieves aggregate consistency within group j is that the shares of the members of group j

---

[2] A downside of the share function approach is that attention must be restricted to non-null equilibria in which $X^j > 0$, and whether a null equilibrium also exists considered separately. Where a null equilibrium is considered it is referred to explicitly, reserving 'Nash equilibrium' for an equilibrium in which some individuals are active.

sum to one. As such, the group-j best response must satisfy

$$\sum_{i \in I^j} s_i^j(X^j; \mathbf{X}^{-j}) = 1. \tag{6}$$

Clearly, the properties of share functions are important in determining the properties of the group best response, elucidating the details of which is turned to next. In applications, it is often straightforward to understand what conditions on the primitives imply share functions are monotonically decreasing in X^j. The following assumption details the conditions on preferences required in the general model.

Assumption 1 *For each individual $i \in I^j$, suppose that u_i^j is continuously differentiable as many times as required, and that:*

a) $\frac{\partial^2 u_i^j}{\partial (x_i^j)^2} < 0$;

b) if $\frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j} < 0$ *then* $\left| \frac{\partial^2 u_i^j}{\partial (x_i^j)^2} \right| > \left| \frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j} \right|$*; and*

c) $\frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j} + \sigma_i^j \left[\frac{\partial^2 u_i^j}{\partial (x_i^j)^2} - \frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j} \right] < 0$ *for all* $\sigma_i^j \in (0, 1]$.

In addition, $\lim_{x_i^j \to \infty} \frac{\partial u_i^j(x_i^j, \cdot; \cdot)}{\partial x_i^j} < 0$.

Thus, payoffs are strictly concave, as previously noted; the substitutability or complementarity of actions within groups must not be too strong; and players will always use a finite action. The following proposition details the properties of share functions; Fig. 1 illustrates.

Proposition 3 *Suppose the preferences of player $i \in I^j$ satisfy Assumption 1. Each player's share function $s_i^j(X^j; \mathbf{X}^{-j})$ is defined for all $X^j \geq \underline{X}_i^j(\mathbf{X}^{-j})$ which is X^j such that $l_i^j(1, X^j; \mathbf{X}^{-j}) = 0$ if this is strictly positive, otherwise the share function is defined for all $X^j > 0$. Define player i's drop-out value as $\bar{X}_i^j(\mathbf{X}^{-j})$ which is X^j such that $l_i^j(0, X^j; \mathbf{X}^{-j}) = 0$ if this exists, or $+\infty$ if it does not. Then the share function has the following properties:*

a) $s_i^j(X^j; \mathbf{X}^{-j}) = 0$ *for all* $X^j \geq \bar{X}_i^j(\mathbf{X}^{-j})$*;*

b) it is continuous and, where it is defined, strictly decreasing in $X^j < \bar{X}_i^j(\mathbf{X}^{-j})$*; and*

c) if $\underline{X}_i^j(\mathbf{X}^{-j}) > 0$ *then* $s_i^j(\underline{X}_i^j(\mathbf{X}^{-j}); \mathbf{X}^{-j}) = 1$, *otherwise* $\lim_{X^j \to 0} s_i^j(X^j; \mathbf{X}^{-j}) = \bar{s}_i^j(\mathbf{X}^{-j})$.

The consistency condition (6) that identifies an equilibrium in the partial game $\mathcal{G}^j(\mathbf{X}^{-j})$ requires X^j to be such that the sum of individual share functions equals one, for then the sum of the individual actions consistent with a Nash equilibrium in which the aggregate action is X^j will sum precisely to X^j. Note that if $\underline{X}_i^j(\mathbf{X}^{-j}) > 0$ for any member of group j then the aggregate share function is defined only where all members' share functions are defined, i.e. for $X^j \geq \max_{i \in I^j}\{\underline{X}_i^j(\mathbf{X}^{-j})\}$. If individual share functions are strictly decreasing in X^j, the aggregation of these will also be strictly decreasing, so under the conditions of Proposition 3 there will be at most one Nash equilibrium. For large enough values of X^j it can be shown that the aggregate share function will take values less than one, so whether a Nash equilibrium exists depends on whether, when X^j is small, the aggregate share function exceeds one. The following proposition makes the conditions for the existence of a unique Nash equilibrium clear.

Proposition 4 *Suppose the preferences of all members of group j satisfy Assumption 1. Then there is a unique Nash equilibrium in the partial game* $\mathcal{G}^j(\mathbf{X}^{-j})$ *in which the aggregate action of the members of group j is* $X^j > 0$ *if either* $\underline{X}_i^j(\mathbf{X}^{-j}) > 0$ *for any* $i \in I^j$, *or* $\sum_{i \in I^j} \bar{s}_i^j(\mathbf{X}^{-j}) > 1$.

Under the conditions stated in the proposition the group best response function $\tilde{X}^j(\mathbf{X}^{-j})$ satisfies

$$\sum_{i \in I^j} s_i^j(\tilde{X}^j(\mathbf{X}^{-j}); \mathbf{X}^{-j}) = 1. \tag{7}$$

The equilibrium action of individual $i \in I^j$ is given by $\tilde{X}^j(\mathbf{X}^{-j}) \sum_{i \in I^j} s_i^j(\tilde{X}^j(\mathbf{X}^{-j}); \mathbf{X}^{-j})$, which is positive if $\tilde{X}^j(\mathbf{X}^{-j}) < \bar{X}_i^j(\mathbf{X}^{-j})$; for some players this inequality may not hold in which case their equilibrium action is zero. If the conditions stated in the proposition are not satisfied then the only Nash equilibrium involves all group members choosing $x_i^j = 0$ and so $\tilde{X}^j(\mathbf{X}^{-j}) \equiv 0$ in these circumstances.[3]

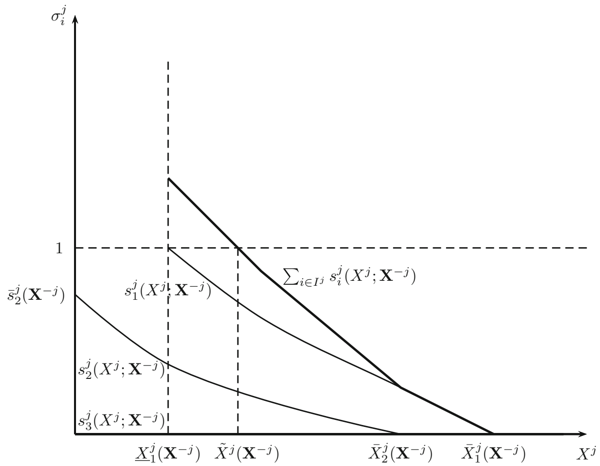


Fig. 1 Illustrating share functions. For individual 1, $\underline{X}_1^j(\mathbf{X}^{-j}) > 0$; for individual 2 $\underline{X}_2^j(\mathbf{X}^{-j}) = 0$ and $\bar{X}_2^j(\mathbf{X}^{-j}) > 0$. For individual 3 $\underline{X}_3^j(\mathbf{X}^{-j}) = \bar{X}_3^j(\mathbf{X}^{-j}) = 0$ and therefore their share function is zero for all $X^j > 0$. The diagram also illustrates the aggregation of share functions for these three players and, assuming they constitute group j, the equilibrium aggregate action in the partial game is identified

[3]The justification for this definition comes from thinking about replacement functions. If $\underline{X}_i^j(\mathbf{X}^{-j}) = 0$ for all $i \in I^j$ then the replacement function is defined for all $X^j \geq 0$, and will take the value zero at $X^j = 0$ (since by definition the replacement value must not exceed X^j).

4 Nash Equilibrium in the Full Game

Having established the existence and uniqueness of Nash equilibrium in group partial games, which has allowed group best response functions $\tilde{X}^j(\mathbf{X}^{-j})$ to be defined, attention now turns to consider mutual consistency of these group best response functions and Nash equilibrium in the full game $\mathcal{G}$. Inter-group consistency of actions occurs with a vector of group aggregates $\mathbf{X}^*$ such that $X^{j*} = \tilde{X}^j(\mathbf{X}^{-j*})$ for all $j \in J$. Finding group aggregate actions that are mutually consistent in the sense of group best responses is analogous to finding mutually consistent individual actions within an N-player game; as with the standard problem, understanding the characteristics of best response functions is important in understanding the features of the equilibrium.

The group structure of the game makes it natural to assume that the members within each group are influenced by the actions of other groups in the same way, captured in the following definition.

Definition The members of group j are 'qualitatively symmetric' if, for any $h, i \in I^j$, $\text{sgn}\{\frac{\partial^2 u_h^j}{\partial x_h^j \partial X^k}\} = \text{sgn}\{\frac{\partial^2 u_i^j}{\partial x_i^j \partial X^k}\}$ and $\text{sgn}\{\frac{\partial u_h^j}{\partial X^k}\} = \text{sgn}\{\frac{\partial u_i^j}{\partial X^k}\}$ for all $k \neq j \in J$.

Formally, only the first of the conditions, that says the marginal payoff of each group member is influenced by the actions of another group in the same direction, is required but it is very natural to also assume the externality from other groups takes the same sign for members of the same group. Note that assuming the members of group j are qualitatively symmetric does not impose that individuals within groups are homogeneous, nor does it restrict the effect of any two different groups on the members of group j to be the same.

The following proposition clarifies the behaviour of group best response functions.

Proposition 5 *Suppose the preferences of all members of all groups satisfy Assumption 1 and that the members of each group are qualitatively symmetric. Then for each $j \in J$ and all $k \neq j \in J$, $\tilde{X}^j(\mathbf{X}^{-j})$ is a continuous function of X^k, and, defining $\bar{\mathcal{X}}^j \subset \mathbb{R}_+^{N-1}$ as the set of values of $\mathbf{X}^{-j}$ where $\tilde{X}^j(\mathbf{X}^{-j}) = 0$,*

$$\text{sgn}\left\{\frac{\partial \tilde{X}^j(\mathbf{X}^{-j})}{\partial X^k}\right\} = \text{sgn}\left\{\frac{\partial^2 u_i^j}{\partial x_i^j \partial X^k}\right\}$$

Taking the sum of these replacement functions (of which a fixed point is sought), if the slope at $X^j \approx 0$ does not exceed 1 (which is intimately related to the condition stated in Proposition 4) then (given share functions are decreasing) it will never exceed 1, and so the only fixed point will be at $X^j = 0$.

for all $\mathbf{X}^{-j} \in \mathbb{R}_+^{N-1} \setminus \bar{\mathcal{X}}^j$. *Moreover, if the strategic effects within group j are stronger than the strategic effects between group j and group k, which requires that* $\left|\frac{\partial^2 u_i^j}{\partial x_i^j \partial X^k}\right| <$ $\left|\sigma_i^j \frac{\partial^2 u_i^j}{\partial (x_i^j)^2} + [1-\sigma_i^j]\frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j}\right|$ *for all* $\sigma_i^j \in [0,1]$, *for all* $i \in I^j$, *then* $\left|\frac{\partial \tilde{X}^j(\mathbf{X}^{-j})}{\partial X^k}\right| < 1$.

Mutual consistency of group best responses requires the identification of a fixed point of the N-dimensional joint group best response function. Since strong assumptions about the differentiability of individuals' payoffs are made, group best responses are continuous functions and therefore the existence of a Nash equilibrium is ensured by Brouwer's fixed point theorem (so long as an assumption of bounded aggregate strategy spaces can also be made). In applications, careful consideration might also be given to whether this fixed point lies in the interior of the aggregate action space, or whether some groups are inactive in equilibrium (i.e. if $\mathbf{X}^{-j} \in \bar{\mathcal{X}}^j$ for any $j \in J$).

It may be possible to say something not only of existence, but also of uniqueness; and to study the comparative static properties of equilibrium: using the approach of group best responses, a comparative static exercise will reveal the effect on group aggregates directly, which are often of primary interest, without having to first deduce the effect on individual equilibrium behaviour that is then aggregated. In applications group best response functions may have clear properties that allow definitive statements about the nature of equilibrium to be made. Whilst it is difficult to draw conclusions in such a general setting as this, the following two statements can be made:

1. If $\tilde{X}^j(\mathbf{X}^{-j})$ is increasing in X^k for all $k \neq j \in J$, then the strategic interaction at the level of groups exhibits complementarities, and the insights from the study of supermodular games (see, for example, Vives 1990) can be used to understand the comparative static properties of the equilibrium group aggregates at the extremal equilibria.[4]
2. If $N = 2$ and the absolute value of the slope of group best response functions for each group (that can be drawn in the space of aggregate group actions) is less than 1, the conditions for which are presented in Proposition 5, the joint group best response will be a contraction and so there will be a unique Nash equilibrium.[5] Whilst there might be many heterogeneous players within groups,

[4]Note, however, that 'group payoff functions' are not defined, so the ideas of supermodular games need only be applied to group best responses. An interesting line of inquiry lies in considering whether, for each group, a payoff function can be defined that, when optimised over the choice of group aggregate (taking the aggregates of other groups as fixed) identifies the same group aggregate as that at the Nash equilibrium within the group. This requires the partial game to be a 'potential game' (Monderer and Shapley 1996), study of which would be an interesting alternative approach to that taken here.

[5]With more than two groups and a desire for uniqueness of equilibrium when the game does not have the features of a nested aggregative game (see below), the approach of Rosen (1965) might be appealed to.

using a multiple aggregate game approach renders the study of the comparative static properties of the equilibrium relatively simple, for the effect of a change in the economic environment on group aggregates in the two-group case will follow from a simple diagrammatic exercise.

5 Nested Aggregative Games

If a strategic interaction has the features of a 'nested aggregative game' more structure can be added to the analysis to draw conclusions about uniqueness of Nash equilibrium in the full game, denoted $\hat{\mathcal{G}}$, even when there are several groups. In such games the members of each group care about their own action, their group's aggregate action, and the *aggregation* of all other groups' aggregate actions, so payoffs can be written $\hat{u}_i^j(x_i^j, X^j; X)$. To analyse this game, the share function approach will be applied twice: once at the level of individuals in groups to replace the fixed point problem of finding consistent actions within groups, as with the analysis so far; and then at the level of groups to replace the fixed point problem of finding consistent aggregate actions between groups.

First fix a value for the aggregate actions of all groups, X, select a group j, and define a 'partial game' $\hat{\mathcal{G}}^j(X)$ in which only the members of group j are considered. The analysis of the partial games is slightly different since whilst the aggregate X is treated as fixed, the influence of players within the group on this aggregate must be accounted for. The aim is to find a group aggregate action that is consistent with a Nash equilibrium in $\hat{\mathcal{G}}$ in which the aggregate of all individuals is X, which means that the group aggregate action must be consistent with the behaviour of members of the group.

Thus, in $\hat{\mathcal{G}}^j(X)$ consider the actions of each member of the group consistent with a Nash equilibrium in which the aggregate of all players is X, and the aggregate of the members of group j is X^j. Share functions that represent consistent individual behaviour in this partial game are denoted $\hat{s}_i^j(X^j; X)$, and take the form $\hat{s}_i^j(X^j; X) = \max\{0, \sigma_i^j\}$ where σ_i^j is the solution to

$$\hat{l}_i^j(\sigma_i^j, X^j; X) \equiv \frac{\mathrm{d}\hat{u}_i^j(\sigma_i^j X^j, X^j; X)}{\mathrm{d}x_i^j} = \frac{\partial \hat{u}_i^j}{\partial x_i^j} + \frac{\partial \hat{u}_i^j}{\partial X^j} + \frac{\partial \hat{u}_i^j}{\partial X} = 0 \tag{8}$$

so long as this does not exceed 1, otherwise the share function is undefined. Consistency of the aggregate action of the members of group j requires individual actions to sum to this aggregate action, or for $\sum_{i \in I^j} \hat{s}_i^j(X^j; X) = 1$. On varying X, this gives the 'group *replacement* function' $\hat{X}^j(X)$, and to find a Nash equilibrium in the game $\hat{\mathcal{G}}$ an aggregate action X must be found that is consistent with the collective behaviour of groups.

Proposition 6 *In a nested aggregative game $\hat{\mathcal{G}}$,* $\mathbf{x}^*$ *is a Nash equilibrium if and only if*

$$\sum_{j \in J} \hat{X}^j(X^*) = X^*.$$

As with individuals within groups, it is often more convenient to work with a group's share of the overall aggregate action, rather than the group aggregate itself. For $X > 0$, the 'group share function' is $\hat{S}^j(X) = \hat{X}^j(X)/X$; letting $\Lambda^j = X^j/X$ the group share function (if it is positive) is that value of Λ^j that satisfies

$$\hat{L}^j(\Lambda^j, X) \equiv \sum_{i \in I^j} \hat{s}_i^j(\Lambda^j X; X) - 1 = 0 \tag{9}$$

so long as the resulting Λ^j does not exceed 1, in which case the group share function is undefined. Accordingly, taking the aggregation of group share functions to be defined only for values of X where the group share function of all groups is defined, there is a Nash equilibrium in the game $\hat{\mathcal{G}}$ with aggregate action $X > 0$ if and only if

$$\sum_{j \in J} \hat{S}^j(X) = 1.$$

The next proposition collects the features of group share functions that allow conclusions about uniqueness of Nash equilibrium to be drawn (in the proof the properties of individual share functions that are relied on to construct the aggregate share functions are also elucidated).

Proposition 7 *Suppose that for each $i \in I^j, j \in J$ utility $\hat{u}_i^j$ is continuously differentiable in each argument as many times as required, and that $\frac{\mathrm{d}\hat{u}_i^j(\sigma_i^j X^j, X^j; X)}{\mathrm{d}x_i^j}$ is strictly decreasing in each of its arguments. Then within each partial game $\hat{\mathcal{G}}^j(X)$ the group share function $\hat{S}^j(X)$ is defined for all $X \geq \underline{\hat{X}}^j$ which is X such that $\hat{L}^j(1, X) = 0$ if this is strictly positive, otherwise it is defined for all $X > 0$ with $\lim_{X \to 0} \hat{S}^j(X) = \hat{\hat{S}}^j$; and it is positive for all $X < \hat{\hat{X}}^j$ which is X such that $\hat{L}^j(0, X) = 0$. Where it is defined and positive, the group share function is strictly decreasing in X. Consequently, there is at most one Nash equilibrium with $X > 0$ in $\hat{\mathcal{G}}$, and if either $\underline{\hat{X}}^j > 0$ for any $j \in J$, or $\sum_{j \in J} \hat{\hat{S}}^j > 0$ if $\underline{\hat{X}}^j = 0$ for all $j \in J$, there is exactly one such Nash equilibrium.*

Thus, in a nested aggregative game the aggregative properties of the game are exploited twice: once at the level of individuals within groups; and once at the level of groups. Deductions concerning the uniqueness of equilibrium can then be made even when there are many groups of heterogeneous players and indeed, once an equilibrium has been identified, whether all groups are active, and whether all individuals within active groups are active, can be understood: in particular, if there is a Nash equilibrium with aggregate X^* and $\hat{\hat{X}}^j \geq X^*$ then group j will be inactive in equilibrium. If a comparative static exercise were to be undertaken for a nested aggregative game, the effect of a change in the economic environment on individual and therefore group share functions must first be understood, and then the effect on the equilibrium aggregate action can be determined, which may in fact be of primary interest; if desired, the effect on the equilibrium aggregate actions of each group can then be considered, along with the effect on individual group members' actions.

6 Applications

6.1 Group Contests

In a standard (Tullock-style) contest each of several individuals chooses the level of 'effort' to exert in contesting a rent, and their success in doing so is determined by the contest success function. In a 'winner-take-all' contest the rent is indivisible and the contest success function determines the probability of a contestant being awarded the rent; hence the contest (imperfectly) discriminates between contestants, giving a higher probability of winning to contestants that exert more effort. In a 'rent-sharing' contest the rent is perfectly divisible and the contest success function determines the share of the rent awarded to each contestant. This discussion considers contests of the latter variety.

If the set of contestants is $\{1, \ldots, i, \ldots, N\}$, the effort chosen by contestant i is $e_i \geq 0$, the aggregate effort of all contestants is E, and $E_{-i} = E - e_i$, then in a simple Tullock contest the contest success function is given by $\frac{e_i}{e_i + E_{-i}}$ (if $E > 0$, otherwise it is $1/N$), and so if R is the contested rent and there is a unit cost of effort each contestant's payoff takes the form $\frac{e_i}{e_i + E_{-i}} R - e_i$. Extensions to this simple model include non-linear costs of effort $c_i(e_i)$; endogenous determination of the rent whereby $R = f(E)$ (Chung 1996); and of course more general contest success functions in which the impact of effort in the contest is given by $p_i(e_i)$ and the contest success function takes the form $\frac{p_i(e_i)}{\sum_{h=1}^{N} p_h(e_h)}$ (see, for example, Cornes et al. 2005). With these extensions, contests capture a multitude of interesting economic environments, so understanding their properties is of upmost importance. There is, of course, a substantial literature on contests, and several contributions have used the techniques of aggregative games to undertake the analysis; it is clear from the contest success function that a contestant's *share* of the aggregate is important, and it was indeed in the study of a 'joint production game'—in which the collective output of individuals is determined by their aggregate input, which is then shared in proportion to those inputs; a simple Tullock contest with an endogenous rent—that the Cornes-Hartley duo first utilised the share function approach (Cornes and Hartley 2000).

While standard contests are appropriate for modelling many economic environments, in some settings individuals are naturally organised into groups and the group plays at least some role in the outcome of the contest. There are, *inter alia*, three interesting scenarios to consider in the context of group contests:

1. In a *collective contest* each individual belongs to a group and decides on a level of effort to contribute to the group. The collective effort of a group then determines the share of the rent received by that group (or the probability of winning the rent in a winner-take-all contest) which then becomes a public good for the group enjoyed equally by all group members irrespective of their initial choice of effort.
2. In an *intra- and inter-group contest* each individual belongs to a group and decides on a level of effort that contributes to the collective effort of the group which determines the share of the rent awarded to that group (as in a collective contest); and a group member's contribution to this collective effort also influences their allocation of the rent within the group.[6]
3. Individuals within groups may be engaged in a contest in which there are *spill-overs between groups*, captured by the size of the rent that each group enjoys itself being influenced by the actions of other groups. Individuals within groups might be engaged in otherwise independent contests, so the actions of members of other groups only influence the rent being contested within each group. Alternatively, groups might be engaged in a contest with each other where the valuation of the rent by each group is influenced by the actions of other groups, which then either becomes a public good for the group members (as in a collective contest), or is contested within the group (as in an intra- and inter-group contest).

6.1.1 Collective Contests

In a collective contest individual i in group j chooses a level of effort x_i^j to contribute to the group. X^j is then the collective effort of the group. The relative effort of group j, X^j/X, determines the share of an exogenously given rent R awarded to group j that becomes a public good for its members. Let v_i^j be individual i's valuation of the rent and $c_i^j(\cdot)$ their cost of effort, which are possibly different for different individuals within each group. Then the payoff to a typical contestant is given by

$$\hat{u}_i^j(x_i^j, X^j; X) = \frac{X^j}{X} v_i^j - c_i^j(x_i^j).$$

[6]Note that an individual's effort choice determines both their contribution to the collective effort in the inter-group contest, and their action in the intra-group contest. This is different to sequential inter- and intra-group conflict, where first individuals in groups secure a rent via their collective action in a contest between groups; and then individuals within each group (or just in the winning group in a winner-take-all contest) seek to appropriate the group's rent with a separate strategic choice (see, for example, Katz and Tokatlidu 1996).

As such, it is clear that collective contests exhibit the features of a nested aggregative game, since payoffs depend only on individual actions, the aggregate action of the group, and the aggregation of all groups' actions.

Such contests capture the essence of collective action; as Konrad (2009, p. 129) notes, "[given the aggregate effort of other groups] the individual effort contributions to the aggregate group effort are contributions to a public good": the quantity of the public good is X^j/X and individual i's valuation of it is $(X^j/X)v_i^j$. It is of course very interesting to understand the effect of collective action on the outcome of the contest, and there is no lack of literature on this subject. Katz et al. (1990) show that when the cost of effort is linear and the valuation of the rent of each member of a group is the same the aggregate effort of each group is uniquely determined and independent of group size, but there is indeterminacy over the split of aggregate effort within the group. In this case the fact that individuals are in groups plays very little role in the outcome, since groups act as though they are one individual, despite the presence of a free-rider problem within groups. Baik (2008) allows for heterogeneous valuations and shows that only the highest-valuation individuals choose positive effort in equilibrium, the remaining individuals free-riding, so consequently there is under-investment in effort by the group. The analysis of this game is very much in the spirit of the idea of partial games since it proceeds by focussing on a group, fixing the actions of the agents in other groups, and considering what is termed the "group [j]-specific equilibrium", which has a straightforward solution given that a single player in each group contributes to collective effort, or if more than one player contributes then those players are necessarily identical.

That all members of a group except those that value the good highest free-ride on the highest valuation members is sensitive to the assumption of linear costs of effort. If costs are convex (but the same for all group members) then all members of a group will contribute to collective effort, as explained by Esteban and Ray (2001) and neatly summarised in Corchón (2007, Sect. 4.2). In these collective contests with convex costs the idea of the "group size paradox"—that free riding is more acute in large groups, meaning smaller groups are more effective—can be explored: however, it is found that whilst individual effort is lower in larger groups the aggregate effort of the group is higher, in contrast to the paradox. This is true where the group see the contested rent as a public good, and even when there is some congestion of it, so long as it is not too strong. Nitzan and Ueda (2014) have extended this literature to allow for members of groups to have both different valuations and different costs, which they do by utilising what has been called here a nested aggregative game approach to derive group share functions and establish consistency of aggregate actions to identify the Nash equilibrium. Being very tractable, this method of analysis allows free riders within groups that make no contribution to group effort to be identified, and the effect of heterogeneity on a group's performance and the contest outcome to be carefully considered.

6.1.2 Intra- and Inter-Group Contests

In an intra- and inter-group contest it is again the relative collective effort of groups that determine their share of the rent (which is taken to have a common value for all group members), but the allocation of that rent share within the group is influenced by the relative effort of group members. In Nitzan (1991) the intra-group allocation is partially determined by the relative effort of group members, with the remaining rent being distributed in an egalitarian way. If α^j is the proportion of the rent that is equally distributed within group j, the payoff to contestant i in group j is given by

$$\hat{u}_i^j(x_i^j, X^j; X) = \left[\alpha^j \frac{1}{N^j} + (1-\alpha^j)\frac{x_i^j}{X^j}\right]\frac{X^j}{X}R - c_i^j(x_i^j),$$

which is again a nested aggregative game.

If $\alpha^j = 0$ for all groups then the X^j in the intra-group contest success function cancels with that in the inter-group contest success function, and the group structure becomes irrelevant as the contest can be seen as a standard Tullock contest with $\sum_{j\in J} N^j$ contestants; hence the characteristics of groups plays no role in the outcome of the contest. If $\alpha^j = 1$ then the case collapses to a contest that is similar to a collective contest in which the value of the contested rent to group j is given by R/N^j. Nitzan (1991) undertook an in-depth analysis of this contest by appealing to the symmetry of contestants within groups by assuming a linear cost of effort and symmetry of sharing rules for groups, showing that as a larger proportion of the rent within groups is allocated based on relative effort, so the collective effort of groups, and consequently the aggregate effort and dissipation of the rent, increases.

By using a nested aggregative game approach, unrealistic symmetry assumptions can be avoided, allowing contestants to have convex costs of effort that can be different. The analysis would proceed by first fixing the aggregate effort of all groups at X and considering consistency of actions among the members of group j. This will define individual share functions, and the value of X^j such that the sum of these share functions is equal to one will give the group-j reaction function, revealing the aggregate effort of group j consistent with a Nash equilibrium in which the aggregate effort of all groups is X. The Nash equilibrium can be found by identifying the level of aggregate effort of all groups that generates consistent group efforts that exactly sum to it, which is where the sum of group share functions is equal to one.

As Konrad (2009) notes, when groups have different sharing rules it is not necessarily the case that all groups will be active in equilibrium, and indeed it will be the case that when contestants within a group have different costs not all contestants will be active. The multiple aggregate game approach, being very tractable in terms of the representation of behaviour consistent with equilibrium, allows for a full understanding of the composition of equilibrium effort to be understood: once group share functions have been aggregated and the equilibrium aggregate effort of all groups found, each group's share function merely needs evaluating at the equilibrium aggregate effort to check whether or not it is zero; and for those groups where it is strictly positive the share functions of individuals within groups can be evaluated to understand which group members are active. The approach holds much hope for understanding the fine details of the equilibrium in group contests, even in quite general settings.

6.1.3 Contests with Group Spill-Overs

An area where multiple aggregate games is likely to make a strong contribution to the analysis of contests in the future is where there are spill-overs between groups' collective effort in terms of the value of the rent being contested by the group, and there is heterogeneity both within and between groups. This can be captured specifying that the value of the rent contested by group j is given by $f^j(X^j, \mathbf{X}^{-j})$. This specification allows the rent to be influenced by the group's own aggregate effort, but not necessarily so. If $\partial f^j/\partial X^k > (<) 0$ then there are positive (negative) externalities between group k and group j. With this rent structure, there might be no other inter-group conflict as contestants within groups contest their group's rent in N otherwise independent group contests; or there may be additional inter-group conflict since the rent, which is valued differently by different groups, is contested between groups as in a collective contest, which could also be coupled with an inter-group contest. The payoff to a typical contestant in the former case would be of the form $u_i^j(x_i^j, X^j; \mathbf{X}^{-j}) = (x_i^j/X^j)f^j(X^j, \mathbf{X}^{-j}) - c_i^j(x_i^j)$, and for the latter case it would be

$$u_i^j(x_i^j, X^j; \mathbf{X}^{-j}) = \left[\alpha^j \frac{1}{N^j} + (1-\alpha^j)\frac{x_i^j}{X^j}\right]\frac{X^j}{X} f^j(X^j, \mathbf{X}^{-j}) - c_i^j(x_i^j),$$

both of which are multiple aggregate games, but not nested aggregative games.

Contests with group spill-overs hold a wealth of interest in terms of applications. In industrial organisation, the framework could be used to capture Cournot competition (which is a simple Tullock contest where effort is output and the rent is endogenously determined as total revenue in the market) between two (or more) groups of sellers who each produce a homogeneous good that acts as a substitute or complement to the other, so the aggregate actions of one group of sellers influence the total revenue that the other group is contesting. In political settings, individuals within political allegiances might contest a rent and the value of this rent is influenced by the actions of competing groups during the campaign. In international trade, groups of traders located in different countries interact both with each other in the home market and, because of trade, the value of this activity will be influenced by the aggregate actions of firms in different locations.

Using the framework of multiple aggregate games, by first resolving the within-group strategic interaction having fixed the behaviour of other groups, and then seeking mutual consistency of behaviour at the level of groups, gives hope for developing an understanding of the features of equilibrium in these as well as other settings to generate new insight concerning these relatively under-explored yet interesting economic environments.

6.2 Bilateral Oligopoly

Bilateral oligopoly is a model of trade in which there is market power on both sides of the market. Given this, a group-based analysis is likely to be fitting as the actions of traders on each side of the market will affect each other, and they will also affect, and be affected by, the actions of traders on the opposite side of the market in the aggregate. One approach to modelling bilateral oligopoly is using a two-commodity version of a Shapley-Shubik strategic market game (Shapley and Shubik 1977) in which one of the goods takes the role of money, and each trader is endowed either with the good or money. This model was originally introduced by Gabszewicz and Michel (1997) and has seen careful study in the literature by Dickson and Hartley (2008), which inspires this discussion, and Amir and Bloch (2009).

Consider an economic environment in which there is a single good g, and money m, that is populated by traders who have preferences that can be represented by utility functions $v_i(g,m)$. Suppose that the set of traders is partitioned into two groups: group 1 contains individuals that are endowed with $e_i^1 > 0$ units of the good but no money, and are called sellers; group 2 contains individuals endowed with $e_i^2 > 0$ units of money but none of the good, and are called buyers. Each seller decides on an *offer* of the good to make to the market $x_i^1 \geq 0$ to be exchanged for money; and each buyer decides on an amount of money to send to the market $x_i^2 \geq 0$ to be exchanged for the good.[7] The market aggregates these offers and bids, and then sellers are awarded a share of the aggregate amount of money sent to the market in proportion to their offer, so receive $\frac{x_i^1}{X^1}X^2$ units of money; and similarly buyers are awarded their proportional share of the aggregate amount of the good in the market, so receive $\frac{x_i^2}{X^2}X^1$ units of the good.[8]

Bilateral oligopoly is therefore a game between two groups where, within each group, individuals engage in a simple Tullock contest in which they choose actions to contest a perfectly divisible prize, the value of which is determined by the aggregation of actions of members of the other group. Payoffs in this game take the form

$$u_i^j(x_i^j, X_{-i}^j; \mathbf{X}^{-j}) = \begin{cases} v_i\left(e_i^1 - x_i^1, \frac{x_i^1}{x_i^1+X_{-i}^1}X^2\right) \text{ if } j=1, \text{or} \\ v_i\left(\frac{x_i^2}{x_i^2+X_{-i}^2}X^1, e_i^2 - x_i^2\right) \text{ if } j=2, \end{cases}$$

[7] For simplicity, it is assumed that endowments are large enough that they will never be constraining and so are ignored in the definition of strategy sets, and in the analysis.

[8] If either $X^1 = 0$ or $X^2 = 0$, no trader receives anything from the market.

and so the game is a multiple (two) aggregate game, but is not a nested aggregative game.

An analysis using the approach outlined here involves first fixing the actions of the buyers and considering the partial game played by the sellers to deduce a function $\tilde{X}^1(X^2)$ that represents the aggregate offers from the sellers consistent with a Nash equilibrium in which the aggregate bid made by the buyers is X^2; and second considering the partial game played by the buyers when the actions of the sellers are considered fixed to deduce the consistent aggregate bid function of the buyers $\tilde{X}^2(X^1)$. When traders' preferences are 'binormal', which requires the (absolute value of) the marginal rate of substitution of the good for money ($\frac{\partial v_i/\partial g}{\partial v_i/\partial m}$) to be decreasing as consumption of the good increases and increasing as consumption of money increases, the share function of every trader is strictly decreasing in the aggregate on their side of the market (Dickson and Hartley 2008; Dickson 2013). Moreover, as Dickson (2013) showed, if for each seller the ratio of the marginal rate of substitution to m is decreasing in m then the consistent aggregate offer function will be increasing in the aggregate bid; and if for each buyer the product of the marginal rate of substitution and g is increasing in g then the consistent aggregate bid function will be increasing in the aggregate offer, as illustrated in Fig. 2.

Under these conditions on preferences there are 'group complementarities', and therefore the ideas of supermodular games applied to group best response functions could be used to discern some comparative static properties of the extremal equilibria. In the illustration there is a single Nash equilibrium, but in this case, since the game has features that mean the group best responses begin from the origin, uniqueness of (non-null) Nash equilibrium cannot be ascertained by appealing to the contraction principle for, if the slope of each group best response is everywhere

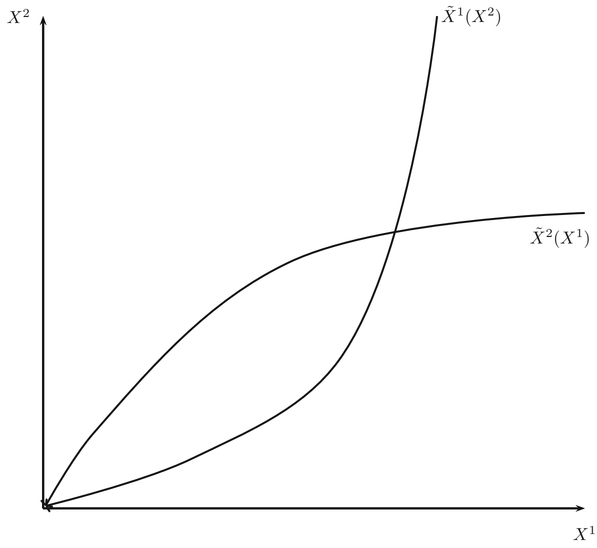


Fig. 2 Consistent aggregate bid and offer functions in bilateral oligopoly

below 1 they will never cross in the interior of aggregate action space. Another possibility is to consider that the slope at the origin exceeds 1 and the group best response functions are concave: in practice many standard utility functions give rise to this property, but deriving an intuitive general statement on preferences is difficult.

To circumvent these issues, Dickson and Hartley (2008) characterised the individual and aggregate behaviour of the two groups consistent with a Nash equilibrium in which the *ratio* of the aggregate money bid to the aggregate amount of commodity offered—which is the price of the good—takes a particular value. For the sellers this is a strategic supply function, and when the consistent aggregate bid of the buyers is divided by the price it is a strategic demand function. Nash equilibria in bilateral oligopoly are identified by the intersections of these strategic versions of Marshallian supply and demand functions, which are monotonic in the expected direction under the stated conditions on preferences and so intersect only once.

This analysis, and in particular study of the uniqueness of Nash equilibrium which is tackled in an environment of heterogeneous traders, is made possible only by taking a multiple aggregate game approach and fixing one side of the market to consider consistent behaviour in the partial game played on the other side of the market. Once the within-group strategic interaction has been resolved and the consistent aggregate behaviour derived, the intersection of strategic supply and demand functions ensures consistency between the sides of the market.

After determining the equilibrium price, equilibrium values of the aggregate offer and bid can be deduced, following which individual traders' strategies can be found, revealing whether there are any inactive traders on either side of the market. Comparative statics are relatively straightforward to undertake to develop an understanding of, for example, the effect of increasing the number of traders on one side of the market, or of increasing the endowment of goods for some sellers, where the effect on the number of sellers that are active in equilibrium might be of relevance.

6.3 Group Public Goods

In an unpublished manuscript, Cornes et al. (2005) consider the provision of public goods within groups when there are spill-overs between groups, capturing the principal of the free-rider problem but where there is also group inter-dependence. If x_i^j is the contribution of player i to the public good of group j then the quantity of the public good provided by group j is X^j and, because of the spill-overs between groups the level of the public good consumed by individuals in group j is given by $X^j + \sum_{k \neq j \in J} \theta^{j,k} X^k$, where $\theta^{j,k}$ is the spill-over parameter capturing how the public good provided by group k influences the members of group j. If there is concordance of interests between group k and j then $\theta^{j,k} > 0$, whilst if their interests are conflicting $\theta^{j,k} < 0$. The consumption of the private good is $m_i^j - c_i^j x_i^j$ where m_i^j

is income and c_i^j the cost of public good provision. The payoff to a typical player is thus

$$u_i^j(x_i^j, X^j; \mathbf{X}^{-j}) = v_i^j(m_i^j - x_i^j, X^j + \sum_{k \neq j \in J} \theta^{j,k} X^k),$$

where $v_i^j(\cdot, \cdot)$ is the player's utility defined over private good consumption and public good consumption. More general formulations of the 'public good production function' for each group could be considered that might account for complementarities between public goods, for example. So long as it is only the aggregate provision of public goods by other groups that matters to an individual, this is a multiple aggregate game.

If the spill-over parameter is common for all groups, i.e. $\theta^{j,k} = \theta$ for all $k \neq j \in J$, for all $j \in J$, then the quantity of the public good consumed by group j can be written $\theta X + (1 - \theta)X^j$. In this case, payoffs are

$$\hat{u}_i^j(x_i^j, X^j, X) = v_i^j(m_i^j - x_i^j, \theta X + (1 - \theta)X^j),$$

and therefore the game is a nested aggregative game.

Cornes et al. (2005) investigate the latter formulation using a replacement function approach. Individual replacement functions $\hat{r}_i^j(X^j; X)$ give the contribution of player i in group j consistent with a Nash equilibrium in which the aggregate contribution of group j is X^j, and the aggregate contribution of all groups is X. They seek group consistency by requiring, for a given X, that $\sum_{i \in I^j} r_i^j(X^j; X) = X^j$ for each $j \in J$ which gives the group-j consistent contribution $\hat{X}^j(X)$, which is a group replacement function. Overall consistency then requires $\sum_{j \in J} \hat{X}^j(X) = X$.

They are able to show that with appropriate restrictions on preferences individual replacement functions are decreasing in X^j and therefore the group-j consistent contribution is unique, so group replacement functions are indeed functions. Moreover, if groups' interests are concordant then individual replacement functions are also decreasing in X which implies that group replacement functions are decreasing in X, so there is a unique value of X where $\sum_{j \in J} \hat{X}^j(X) = X$ and so a unique Nash equilibrium. If group interests are conflicting then group replacement functions are increasing in X, but so long as the conflict is not too strong the function $\sum_{j \in J} \hat{X}^j(X)$ will be a contraction (its slope will be less than one) with a unique fixed point and therefore a unique Nash equilibrium. Given uniqueness of equilibrium, found under quite general conditions, understanding the comparative static properties of equilibrium is a relatively straightforward task.

Cornes et al. (2005) suggest an extension to the case where the good generated by contributions of individuals does not become a pure nonexcludable good, but is distributed among group members according to some sharing rule. This could be captured by considering that the good becomes a private good and is shared among group members according to the rule $[\alpha \frac{1}{N^j} + \frac{x_i^j}{X^j}]$, as in Nitzan (1991). This preserves the nested aggregate nature of the game, and is consistent with the idea of contests between groups where the size of the contested rent is influenced by the actions of other groups, introduced previously, and so could be analysed using the tools of multiple aggregate games.

7 Concluding Remarks

This contribution considers the theory of multiple aggregate games, in which individuals are organised into groups and there are both within-group and between-group strategic tensions that can be very different in nature. This general framework captures environments in which there is a collective element to individual actions within groups, and externalities between groups' aggregate actions. The within-group strategic interaction is assumed to have the features of an aggregative game, and this is exploited in the method proposed to analyse these games, which follows a two step procedure: first, the intra-group strategic interaction is resolved through study of group 'partial games' to derive group best responses; then the inter-group game is analysed by considering mutual consistency of these group best responses at the level of group aggregates. If 'aggregativeness' also pervades the between-group interaction, as in a nested aggregative game, then rather than using group best responses, group replacement functions can be derived that have a much simpler consistency requirement to identify equilibria.

Exploiting the aggregative properties of games makes for a very tractable analysis since the structure of the game is used to reduce the dimensionality of the problem. Replacement (or share) functions often have very clear properties in terms of their monotonicity and where they drop to zero, that are preserved when they are aggregated. Establishing existence and uniqueness of Nash equilibrium; understanding which players are active in equilibrium; and undertaking a comparative static analysis (that may involve adding players) are all relatively straightforward tasks.

By using this method within groups and, if appropriate, also between groups, the analysis of multiple aggregate games becomes much less daunting even with heterogeneity of players in groups, and heterogeneity between groups. In applications, this permits a rather general analysis to be undertaken that has the scope to answer many interesting questions that might be posed, particularly related to the effect of heterogeneity within and between groups. Some applications that have been considered in the literature—collective contests; group provision of public goods; bilateral oligopoly—have been discussed, and others speculated upon. Some of these, as well as many others, fit within a model of group contests with group spill-overs, careful study of which seems to be a fruitful direction for future research. I hope that the exposition of multiple aggregate games presented here is useful in pursuing this and other lines of research, and I also hope that I have done justice to the ideas that Richard and I discussed.

Appendix: Proofs

Proof of Proposition 1 The proof is by simple definition chasing. If $\mathbf{x}^*$ is a Nash equilibrium in $\mathcal{G}$, then by definition $x_i^{j*} = \tilde{x}_i^j(\mathbf{X}^{-j*})$ for all $i \in I^j, j \in J$. But then for each $j \in J$, $X^{j*} = \sum_{i \in I^j} \tilde{x}_i^j(\mathbf{X}^{-j*}) = \tilde{X}^j(\mathbf{X}^{-j*})$. Conversely, if $X^{j*} = \tilde{X}^j(\mathbf{X}^{-j*})$ for all $j \in J$, then by definition $\mathbf{x}^{j*}$ where $x_i^{j*} = \tilde{x}_i^j(\mathbf{X}^{-j*})$ for each $i \in I^j$ is a Nash equilibrium in $\mathcal{G}^j(\mathbf{X}^{-j*})$ for all $j \in J$; it then follows that $\mathbf{x}^*$ is a Nash equilibrium. □

Proof of Proposition 2 The proof is again by definition chasing. First, if $\mathbf{x}^{j*}$ is a Nash equilibrium then $x_i^{j*} = b_i^j(X_{-i}^{j*}; \mathbf{X}^{-j})$ for all $i \in I^j$. This implies $x_i^{j*} = b_i^j(X^{j*} - x_i^{j*}; \mathbf{X}^{-j})$, and so by definition $x_i^{j*} = r_i^j(X^{j*}; \mathbf{X}^{-j})$ for all $i \in I^j$; and therefore $\sum_{i \in I^j} r_i^j(X^{j*}; \mathbf{X}^{-j}) = X^{j*}$. To prove necessity, suppose $\sum_{i \in I^j} r_i^j(X^{j*}; \mathbf{X}^{-j}) = X^{j*}$ and consider the strategy $x_i^{j*} = r_i^j(X^{j*}; \mathbf{X}^{-j})$. By definition of the replacement function, $x_i^{j*} = b_i^j(X^{j*} - x_i^{j*}; \mathbf{X}^{-j})$, and since $X^{j*} = \sum_{i \in I^j} r_i^j(X^{j*}; \mathbf{X}^{-j})$ it follows that $X^{j*} - x_i^{j*} = \sum_{h \neq i \in I^j} r_h^j(X^{j*}; \mathbf{X}^{-j}) = X_{-i}^{j*}$. As such, $x_i^{j*} = b_i^j(X_{-i}^{j*}; \mathbf{X}^{-j})$ for all $i \in I^j$, so $\mathbf{x}^{j*}$ is a Nash equilibrium in $\mathcal{G}^j(\mathbf{X}^{-j})$. □

Proof of Proposition 3 A player's share function is the value of σ_i^j that makes l_i^j, defined in (5), equal to 0—however, if this is below zero the share function is defined as zero; and if it is above 1 the share function is undefined. First, note that under Assumption 1

$$\frac{\partial l_i^j}{\partial \sigma_i^j} = X^j \left[\frac{\partial^2 u_i^j}{\partial (x_i^j)^2} + \frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j} \right] < 0 \text{ and}$$

$$\frac{\partial l_i^j}{\partial X^j} = \sigma_i^j \frac{\partial^2 u_i^j}{\partial (x_i^j)^2} + [1 - \sigma_i^j] \frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j} < 0.$$

The first inequality implies there is at most one $\sigma_i^j \in [0, 1]$ where $l_i^j = 0$ so the share function is indeed a function. Continuity of this function, where it is defined, follows from l_i^j varying continuously in all its arguments by virtue of the assumed differentiability of utility functions.

If $\underline{X}_i^j(\mathbf{X}^{-j}) > 0$, by definition, $l_i^j(1, \underline{X}_i^j(\mathbf{X}^{-j}); \mathbf{X}^{-j}) = 0$, so $s_i^j(\underline{X}_i^j(\mathbf{X}^{-j}), \mathbf{X}^{-j}) = 1$ and the monotonicity properties just stated imply that for all $X^j < \underline{X}_i^j(\mathbf{X}^{-j})$, $l_i^j > 0$ for all $\sigma_i^j \leq 1$, and therefore the share function is undefined. In addition, again by definition, if $\bar{X}_i^j(\mathbf{X}^{-j}) < \infty$ then $l_i^j(0, \bar{X}_i^j(\mathbf{X}^{-j}); \mathbf{X}^{-j}) = 0$ so $s_i^j(\bar{X}_i^j(\mathbf{X}^{-j}); \mathbf{X}^{-j}) = 0$ and the monotonicity properties of l_i^j imply that for all $X^j > \bar{X}_i^j(\mathbf{X}^{-j})$, $l_i^j = 0$ only when $\sigma_i^j < 0$, and therefore by definition $s_i^j(X^j; \mathbf{X}^{-j}) = 0$.

Where it is positive and defined, implicit differentiation of the first-order condition that defines the share function, (5), gives

$$\frac{\partial s_i^j}{\partial X^j} = -\frac{\frac{\partial l_i^j}{\partial X^j}}{\frac{\partial l_i^j}{\partial \sigma_i^j}} < 0,$$

where the inequality follows from the deductions above.

If $\underline{X}_i^j(\mathbf{X}^{-j}) = 0$ the share function is defined for all $X^j > 0$ where it takes values in the compact set $[0, 1]$, so (taking subsequences if necessary) the limit as $X^j \to 0$ will exist, which is denoted $\bar{s}_i^j(\mathbf{X}^{-j})$. □

Proof of Proposition 4 If $\bar{X}_i^j(\mathbf{X}^{-j}) < \infty$ the share function of individual $i \in I^j$ is equal to zero for all $X^j \geq \bar{X}_i^j(\mathbf{X}^{-j})$. If not, then since it is assumed that $\lim_{x_i^j \to \infty} \frac{\partial u_i^j(x_i^j, \cdot; \cdot)}{\partial x_i^j} < 0$ the first-order condition (5) can hold as $X^j \to \infty$ only if $\lim_{X^j \to \infty} \sigma_i^j X^j < \infty$ which requires $\sigma_i^j \to 0$, implying the share function vanishes in the large X^j limit. This implies there is an $\bar{x}^j(\mathbf{X}^{-j})$ such that $\sum_{i \in I^j} s_i^j(X^j; \mathbf{X}^{-j}) < 1$ for all $X^j > \bar{x}^j(\mathbf{X}^{-j})$. The function $\sum_{i \in I^j} s_i^j(X^j; \mathbf{X}^{-j})$ is continuous and strictly decreasing in X^j for all $\max_{i \in I^j}\{\underline{X}_i^j(\mathbf{X}^{-j})\} < X^j < \bar{x}^j(\mathbf{X}^{-j})$, and is therefore equal to 1 for at most one value of X^j. If $\underline{X}_i^j(\mathbf{X}^{-j}) > 0$ for any $i \in I^j$ then $\sum_{i \in I^j} s_i^j(\max_{i \in I^j}\{\underline{X}_i^j(\mathbf{X}^{-j})\}, \mathbf{X}^{-j}) \geq 1$ and so there is a unique value of X^j where $\sum_{i \in I^j} s_i^j(X^j, \mathbf{X}^{-j}) = 1$. If $\underline{X}_i^j(\mathbf{X}^{-j}) = 0$ for all $i \in I^j$ then the aggregate share function is defined for all $X^j > 0$, with $\lim_{X^j \to 0} \sum_{i \in I^j} s_i^j(X^j, \mathbf{X}^{-j}) = \sum_{i \in I^j} \bar{s}_i^j(\mathbf{X}^{-j})$. As such, the existence of a (unique) Nash equilibrium requires $\sum_{i \in I^j} \bar{s}_i^j(\mathbf{X}^{-j}) > 1$. □

Proof of Proposition 5 For $\mathbf{X}^{-j} \in \mathbb{R}_+^{N-1} \setminus \bar{\mathcal{X}}^j$, the group best response is implicitly defined by (7). Continuity of the group best response follows from continuity of individual share functions in each of its arguments, which follows from the assumed differentiability of utility functions. With apology[9] implicit differentiation of (7) gives

$$\frac{\partial \tilde{X}^j(\mathbf{X}^{-j})}{\partial X^k} = -\frac{\sum_{i \in I^j} \frac{\partial s_i^j}{\partial X^k}}{\sum_{i \in I^j} \frac{\partial s_i^j}{\partial X^j}}. \tag{10}$$

[9]Whilst individual share functions very smoothly in their arguments, the aggregation of these within a group, whilst continuous, does not necessarily vary in a smooth way, in particular in a neighborhood of a group member's 'dropout value' $\bar{X}_i^j(\mathbf{X}^{-j})$. As such, implicit differentiation should not be used at these points on the domain but, with apology, it is given its intuitive merit. In a neighborhood of any $\bar{X}_i^j(\mathbf{X}^{-j})$ the derived derivatives do not hold and indeed should not be defined; the monotonicity properties can nevertheless be proved for these regions of the domain by a contradictory argument (details omitted).

Where an individual's share function is positive, recall that it is defined by the first-order condition $l_i^j(\sigma_i^j, X^j; \mathbf{X}^{-j}) = 0$ as in (5). As deduced previously,

$$\frac{\partial s_i^j}{\partial X^j} = -\frac{\frac{\partial l_i^j}{\partial X^j}}{\frac{\partial l_i^j}{\partial \sigma_i^j}} = \frac{\sigma_i^j \frac{\partial^2 u_i^j}{\partial (x_i^j)^2} + [1-\sigma_i^j]\frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j}}{X^j\left[\frac{\partial^2 u_i^j}{\partial (x_i^j)^2} - \frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j}\right]} < 0$$

under Assumption 1. In addition,

$$\frac{\partial s_i^j}{\partial X^k} = -\frac{\frac{\partial l_i^j}{\partial X^k}}{\frac{\partial l_i^j}{\partial \sigma_i^j}} = -\frac{\frac{\partial^2 u_i^j}{\partial x_i^j \partial X^k}}{X^j\left[\frac{\partial^2 u_i^j}{\partial (x_i^j)^2} - \frac{\partial^2 u_i^j}{\partial x_i^j \partial X_{-i}^j}\right]}.$$

Since the denominator is negative under Assumption 1, $\mathrm{sgn}\{\frac{\partial s_i^j}{\partial X^k}\} = \mathrm{sgn}\{\frac{\partial^2 u_i^j}{\partial x_i^j \partial X^k}\}$. As such, since group members are qualitatively symmetric, it follows that $\mathrm{sgn}\{\frac{\partial \tilde{X}^j(\mathbf{X}^{-j})}{\partial X^k}\} = \mathrm{sgn}\{\frac{\partial^2 u_i^j}{\partial x_i^j \partial X^k}\}$, as stated.

$\left|\frac{\partial \tilde{X}^j(\mathbf{X}^{-j})}{\partial X^k}\right| < 1$ if the numerator in (10) is less than the denominator, a sufficient (but by no means necessary) condition for which is $\left|\frac{\partial s_i^j}{\partial X^k}\right| < \left|\frac{\partial s_i^j}{\partial X^j}\right|$ for all $i \in I^j$, which is implied by the inequality in the proposition. □

Proof of Proposition 6 If $\mathbf{x}^*$ is a Nash equilibrium, then by definition of share functions $x_i^{j*} = X^{j*}\hat{s}_i^j(X^{j*}; X^*)$ for all $i \in I^j, j \in J$. As such, $\sum_{i\in I^j} \hat{s}_i^j(X^{j*}; X^*) = 1$ and therefore $X^{j*} = \hat{X}^j(X^*)$ for all $j \in J$, implying $X^* = \sum_{j\in J} \hat{X}^j(X^*)$. For necessity of the condition, define a player's best response in a nested aggregative game as $b_i^j(X_{-i}^j; X^{-j})$. Consider the strategy $x_i^{j*} = \hat{X}^j(X^*)\hat{s}_i^j(X^{j*}; X^*)$. By definition of share functions and the consistency of $\hat{X}^j(X^*)$ within group j (which implies $\hat{X}^j(X^*) - x_i^{j*} = X_{-i}^{j*}$), $x_i^{j*} = b_i^j(X_{-i}^{j*}; X^* - \hat{X}^j(X^*))$. When $X^* = \sum_{j\in J} \hat{X}^j(X^*)$, it follows that $X^* - \hat{X}^j(X^*) = X^{-j*}$, and therefore $x_i^{j*} = b_i^j(X_{-i}^{j*}; X^{-j*})$ for all $i \in I^j, j \in J$, giving the conclusion that $\mathbf{x}^*$ is a Nash equilibrium. □

Proof of Proposition 7 The properties of individual share functions in $\hat{\mathcal{G}}^j(X)$ are first deduced. The conditions stated on preferences are equivalent to assuming

$$\frac{\partial \hat{l}_i^j}{\partial \sigma_i^j} = X^j\left[\frac{\partial^2 \hat{u}_i^j}{\partial (x_i^j)^2} + \frac{\partial^2 \hat{u}_i^j}{\partial x_i^j \partial X^j} + \frac{\partial^2 \hat{u}_i^j}{\partial x_i^j \partial X}\right] < 0,$$

$$\frac{\partial \hat{l}_i^j}{\partial X^j} = \sigma_i^j \frac{\partial^2 \hat{u}_i^j}{\partial (x_i^j)^2} + \frac{\partial^2 \hat{u}_i^j}{\partial x_i^j \partial X^j} + \sigma_i^j \frac{\partial^2 \hat{u}_i^j}{\partial x_i^j \partial X^j} + \frac{\partial^2 \hat{u}_i^j}{\partial (X^j)^2} + \sigma_i^j \frac{\partial^2 \hat{u}_i^j}{\partial x_i^j \partial X} + \frac{\partial^2 \hat{u}_i^j}{\partial X^j \partial X} < 0, \text{ and}$$

$$\frac{\partial \hat{l}_i^j}{\partial X} = \frac{\partial^2 \hat{u}_i^j}{\partial x_i^j \partial X} + \frac{\partial^2 \hat{u}_i^j}{\partial X^j \partial X} + \frac{\partial^2 \hat{u}_i^j}{\partial (X)^2} < 0.$$

Under the first two conditions (as previously) two thresholds $\underline{\hat{X}}_i^j(X)$ (which is $X^j > 0$ such that $\hat{l}_i^j(1, X^j; X) = 1$) and $\hat{\hat{X}}_i^j(X)$ (which is X^j such that $\hat{l}_i^j(0, X^j; X) = 1$ if such an X^j exists, otherwise it is defined as $+\infty$) can be defined, between which the share function is defined and takes positive values, and where

$$\frac{\partial \hat{s}_i^j}{\partial X^j} = -\frac{\frac{\partial \hat{l}_i^j}{\partial X^j}}{\frac{\partial \hat{l}_i^j}{\partial \sigma_i^j}} < 0.$$

If $\underline{\hat{X}}_i^j(X)$ as defined above does not exist then the share function is defined for all $X^j > 0$ with $\lim_{X^j \to 0} \hat{s}_i^j(X^j; X) = \hat{\bar{s}}_i^j(X)$.

As before, the aggregation of individual share functions is taken to be defined only for values of X^j where all group members' share functions are defined. Noting that share functions are either equal to zero for large enough X^j, or are vanishing in the large X^j limit, if either $\underline{\hat{X}}_i^j(X) > 0$ for any $i \in I^j$, or $\sum_{i \in I^j} \hat{\bar{s}}_i^j(X) > 1$ then there is a single consistent aggregate action $\hat{X}^j(X)$ in $\hat{\mathcal{G}}^j$ which is such that $\sum_{i \in I^j} \hat{s}_i^j(\hat{X}^j(X); X) = 1$. If this is not the case then $\hat{X}^j(X) = 0$.

Consider now varying X to change the partial game played by group j. Group j's share of the total aggregate is $\hat{S}^j(X) = \hat{X}^j(X)/X$, defined by (9) if the resulting share is between 0 and 1. Note that

$$\frac{\partial \hat{L}^j}{\partial \Lambda^j} = X \sum_{i \in I^j} \frac{\partial \hat{s}_i^j}{\partial X^j} \text{and}$$

$$\frac{\partial \hat{L}^j}{\partial X} = \sum_{i \in I^j} \Lambda^j \frac{\partial \hat{s}_i^j}{\partial X^j} + \frac{\partial \hat{s}_i^j}{\partial X}.$$

Now,

$$\frac{\partial \hat{s}_i^j}{\partial X} = -\frac{\frac{\partial \hat{l}_i^j}{\partial X}}{\frac{\partial \hat{l}_i^j}{\partial \sigma_i^j}} < 0,$$

which, combined with the monotonicity of share functions with respect to group aggregate, implies both of the expressions above are negative. Given this, the thresholds and the monotonicity of group share functions stated in the proposition can be derived analogously to the case within groups, so the details are omitted. Aggregate share functions are either equal to zero for $X > \hat{\hat{X}}^j$ or, if this is not finite, vanish in the large X limit—this follows by recalling that individual share functions vanish in the large X^j limit, so as $X \to \infty$ (9) can hold only if $\lim_{X \to \infty} \Lambda^j X < \infty$ which requires $\Lambda^j \to 0$. Given this, if either $\underline{\hat{X}}^j > 0$ for any $j \in J$, or $\sum_{j \in J} \hat{\bar{S}}^j > 0$ if

$\underline{\hat{X}}^j = 0$ for all $j \in J$, the aggregate share function will exceed one for small enough X and since it is strictly decreasing in X will be equal to one at exactly one value of X, so consequently there is a unique Nash equilibrium. □

References

Acemoglu, D., & Jensen, M. K. (2013). Aggregate comparative statics. *Games and Economic Behavior, 81*, 27–49.

Amir, R., & Bloch, F. (2009). Comparative statics in a simple class of strategic market games. *Games and Economic Behavior, 65*(1), 7–24.

Baik, K. H. (2008). Contests with group-specific public-good prizes. *Social Choice and Welfare, 30*(1), 103–117.

Chung, T.-Y. (1996). Rent-seeking contest when the prize increases with aggregate efforts. *Public Choice, 87*(1–2), 55–66.

Corchon, L. C. (1994). Comparative statics for aggregative games the strong concavity case. *Mathematical Social Sciences, 28*(3), 151–165.

Corchón, L. C. (2007). The theory of contests: A survey. *Review of Economic Design, 11*(2), 69–100.

Cornes, R., & Hartley, R. (2000). Joint production games and share functions. Mimeo.

Cornes, R., & Hartley, R. (2007). Aggregative public good games. *Journal of Public Economic Theory, 9*(2), 201–219.

Cornes, R., & Hartley, R. (2012). Fully aggregative games. *Economics Letters, 116*(3), 631–633.

Cornes, R., Hartley, R., & Nelson, D. (2005). Groups with intersecting interests. Mimeo.

Cornes, R. C., Hartley, R., & Tamura, Y. (2010). Two-aggregate games: A method of analysis and an example. Mimeo.

Dickson, A. (2013). On Cobb-Douglas preferences in bilateral oligopoly. *Recherches économiques de Louvain, 79*(4), 89–110.

Dickson, A., & Hartley, R. (2008). The strategic Marshallian cross. *Games and Economic Behavior, 64*(2), 514–532.

Esteban, J., & Ray, D. (2001). Collective action and the group size paradox. *American Political Science Review, 95*, 663–672. Cambridge University Press.

Gabszewicz, J., & Michel, P. (1997). Oligopoly equilibrium in exchange economies. In B. Eaton & R. Harris (Eds.), *Trade, technology and economics: Essays in honor of Richard G. Lipsey* (pp. 217–240). Cheltenham: Elgar.

Jensen, M. K. (2010). Aggregative games and best-reply potentials. *Economic Theory, 43*(1), 45–66.

Katz, E., Nitzan, S., & Rosenberg, J. (1990). Rent-seeking for pure public goods. *Public Choice, 65*(1), 49–60.

Katz, E., & Tokatlidu, J. (1996). Group competition for rents. *European Journal of Political Economy, 12*(4), 599–607.

Kolmar, M., & Rommeswinkel, H. (2013). Contests with group-specific public goods and complementarities in efforts. *Journal of Economic Behavior and Organization, 89*, 9–22.

Konrad, K. A. (2009). *Strategy and dynamics in contests*. Oxford: Oxford University Press.

Monderer, D., & Shapley, L. S. (1996). Potential games. *Games and Economic Behavior, 14*(1), 124–143.

Nitzan, S. (1991). Collective rent dissipation. *The Economic Journal, 101*(409), 1522–1534.

Nitzan, S., & Ueda, K. (2014). Intra-group heterogeneity in collective contests. *Social Choice and Welfare, 43*(1), 219–238.

Novshek, W. (1985). On the existence of Cournot equilibrium. *The Review of Economic Studies, 52*(1), 85–98.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica, 33*(3), 520–534.

Selten, R. (1970). *Preispolitik der Mehrproduktenunternehmung in der statischen Theorie*. Berlin: Springer.

Shapley, L., & Shubik, M. (1977). Trade using one commodity as a means of payment. *The Journal of Political Economy, 85*(5), 937–968.

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics, 19*(3), 305–321.

Wolfstetter, E. (1999). *Topics in microeconomics: Industrial organization, auctions, and incentives*. Cambridge: Cambridge University Press.

Strategic Coalition Formation in Global Public Good Provision

Wolfgang Buchholz and Michael Eichenseer

1 Introduction

Dating back to Wicksell (1896) and Lindahl (1919) the analysis of public good provision has been a cornerstone in public economic theory, which provided one of the main economic rationales for governmental activities in a market economy. In the last few decades, attention has shifted from public goods that are allocated by a central authority to public goods (as, e.g., charities), which are instead provided through voluntary "private" contributions by the agents involved. In this context, a key insight has been that private provision of public goods (in the ensuing Nash equilibrium) usually does not lead to an efficient solution but rather to some "underprovision" of the public good as compared to Pareto optimal levels [see the path-breaking works of Cornes and Sandler (1986/1996), and Bergstrom et al. (1986)]. This inefficiency is especially pertinent for international public goods (as disease eradication, military defense and counterterrorism, as well as transboundary pollution externalities), which—in a parallel, but more or less accidental development to that of public good theory—have increasingly got public attention and meanwhile have become a central issue of international politics [see, e.g., Sandler (2004) and Peinhardt and Sandler (2015)].

In order to overcome the potentially fatal underprovision of international public goods, an obvious idea seems to be to transfer the top-down approach being applied for public good allocation at the national level to the international sphere. In the

W. Buchholz (✉)
Institute of Economics and Econometrics, University of Regensburg, Regensburg, Germany
e-mail: Wolfgang.Buchholz@ur.de

M. Eichenseer
Department of Economics, University of Regensburg, 93040 Regensburg, Germany
e-mail: Michael.Eichenseer@ur.de

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_4

Lindahlian tradition this in particular means that—analogous to a taxation scheme—a fair burden sharing rule has to be established according to which the country-specific contributions to the public good are determined. It therefore does not come as a surprise that initially such a route has been taken in international negotiations on climate protection, which by now is considered as the most important global public good. But experience over many rounds of the "Conferences of the Parties" since the 1990s has demonstrated some quite foreseeable flaws of this top-down approach, which has inhibited global climate cooperation at the needed scale and which, in particular, has made the Kyoto Protocol as the basic international climate agreement only little effective. So, due to the absence of a central authority with coercive power and alternative sanction mechanisms, the pervasive free-rider problem could not be solved, and the enormous difficulties to reconcile divergent notions on equitable effort-sharing have repeatedly turned up in climate negotiations.

In face of these deficiencies global climate policy has changed its course to a bottom-up approach, which recently found its expression in the "Paris Agreement on Climate Change" concluded in December 2015 (e.g., Sandler 2016; Chen and Zeckhauser 2016). The expectation, or realistically rather hope, which is set on this agreement is that—beyond creating more transparency—the pre-announced greenhouse gas abatement measures of some countries should induce higher abatement efforts of other countries and thus trigger a dynamic process towards a more effective global climate protection.

For public good theory this change of paradigm in global climate policy provides some novel challenges as, e.g., the need of a deeper understanding of the role which the demonstration of good intentions by some agents can play for the voluntary provision of public goods [see Buchholz and Sandler (2016)]. In this paper, we will deal with a special one among these issues, i.e., how coalition formation by one group of countries influences the coalition formation decision of another group. It crucially depends on this incentive effect whether it can be expected that cooperation within one group stimulates cooperation by another group, which clearly is relevant for assessing the prospects of success of the bottom-up approach. In particular, we will explore in a two-stage game how the size of the two potential coalitions affect their decision to partial cooperation within their group. As in the more specific analysis by Hattori (2015) we determine the Nash equilibria at the first coalition formation stage of the entire two-stage game, whose second stage is given by a standard model of voluntary public good provision as in Cornes and Sandler (1986/1996). The mirror image of coalition formation is strategic decentralization as considered by Eckert (2003), Buchholz et al. (2013) and Foucart and Wan (2016) who apply a more special framework. The method which is applied for our analysis will be the Aggregative Game Approach (AGA), whose development owes a lot to Richard Cornes' remarkable ingenuity and through which the theoretical analysis of public good provision has been facilitated substantially [see Cornes and Hartley (2007), and with specific application to issues in environmental economics Cornes (2016)].

The paper will be organized as follows: After presenting the theoretical framework in Sect. 2, we describe in Sects. 3 and 4 which Nash equilibria will arise at stage 2, when either one group or both groups have decided to cooperate at stage 1. In Sect. 5 we then analyze a group's decision to form a coalition given that the members of the other group either act in isolation or are cooperating themselves. In this context we will—unlike Buchholz et al. (1998)—not consider utility effects of small variations of a coalition's public good contribution but instead assume unmitigated utility maximization of a coalition as collective Nash reaction to the public good contributions of the other group [as in Buchholz et al. (2014), and Vicary (2016)]. Therefore, corner solutions in which one of the group does not make any public good contribution at all, are of particular importance in the context of our paper. Based on this we explore in Sect. 6 the properties of the Nash equilibria at the coalition formation stage where the groups anticipate the outcome at the second contribution stage, thus characterizing the subgame-perfect equilibria for the entire two-stage game. In Sect. 6 we also briefly consider a version of the game in which one of the two groups acts as the first mover at the coalition formation stage. These results are illustrated by specific examples with Cobb-Douglas preferences in Sect. 7. In Sect. 8 we conclude.

2 The Framework

There are two groups of otherwise identical countries K and M of size $k \geq 2$ and $m \geq 2$, respectively. Each country i is characterized by its initial private good endowment w and its utility function $u(x_i, G)$, where x_i denotes agent i's level of private consumption and G indicates public good supply. Each utility function is assumed to have the standard properties, i.e., it is twice continuously differentiable, quasi-concave, and strictly monotone increasing in both variables. Moreover, both goods are assumed to be non-inferior.

The main ingredient of the AGA, which in the following is used to characterize equilibria, are (income) *expansion paths*, which are well-known from standard household theory. For any marginal rate of substitution $mrs = \rho$ between the private and the public good the associated expansion path is given by $e(G, \rho)$, which is a well-defined and strictly monotone increasing (and differentiable) function of G. In x_i-G-space, such an expansion path connects all points (x_i, G) at which country i's indifference curves have slope $-\rho$ so that $\rho = \frac{\partial u/\partial x_i}{\partial u/\partial G}(x_i, G)$ holds. In order to avoid the tedious treatment of sub-cases, we assume $e(0, \rho) = 0$ and $\lim_{G\to\infty} e(G, \rho) = \infty$, which, e.g., results when preferences are of the Cobb-Douglas type.

As in Andreoni (1988) and Andreoni and McGuire (1993) let $\overline{G}_\rho$ be that level of public good supply for which the condition $e(\overline{G}_\rho, \alpha) = w$ is fulfilled. Convexity of indifference curves implies $e(G, \rho_2) < e(G, \rho_1)$ if $\rho_2 > \rho_1$ and thus $\overline{G}_{\rho_2} > \overline{G}_{\rho_1}$ (see Fig. 1).

The public good is produced by a summation technology. We assume that all countries have the same marginal rate of transformation between the private and the public good, which is normalized to one. If a country i's public good contribution

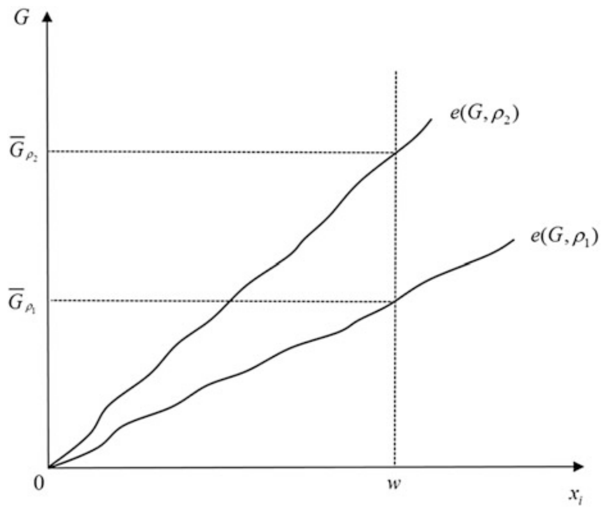


Fig. 1 Expansion paths

is $g_i = w - x_i \geq 0$ aggregate public good supply then becomes $G = \sum_{i=1}^{n} g_i$ where $n := k + m$. An allocation $(x_1, \ldots, x_n, G)$ thus is feasible if and only if

$$G + \sum_{i=1}^{n} x_i = nw \tag{1}$$

holds.

Public good supply in the standard Nash equilibrium $E^N(n)$, in which all n countries act non-cooperatively, is denoted by $G^N(n)$, which is given by the condition

$$G^N(n) + ne(G^N(n), 1) = nw. \tag{2}$$

The characterization of the Nash equilibrium as provided by (2) is based on the feasibility constraint (1) and on the fact that a country, which actively contributes to the public good, only is in an equilibrium position if its marginal rate of substitution coincides with the marginal rate of transformation $mrt = 1$, i.e., if its position is on the expansion path $e(G, 1)$. Since expansion paths are upward sloping, it directly follows from (2) that $G^N(n) < \overline{G}_1$.

The Nash equilibrium $E^N(n)$ only depends on the total number $n = k + m$ of countries, and it is symmetric and interior, i.e., each country makes a strictly positive contribution to the public good. In $E^N(n)$ each country's private consumption is

$x^N(n) = e(G^N(n), 1)$, so that a country's utility in the standard Nash equilibrium is $u^N(n) = u\big(e(G^N(n), 1), G^N(n)\big)$.

In the Cobb-Douglas case with the utility function $u(x_i, G) = x_i^\alpha G$, which will be used as an illustration throughout the paper, we have $e(G, 1) = \alpha G$ so that condition (2) gives $G^N(n) = \frac{nw}{1+n\alpha}$ and thus $x^N(n) = \frac{n\alpha w}{1+n\alpha}$ and $u^N(n) = \alpha^\alpha\left(\frac{nw}{1+n\alpha}\right)^{1+\alpha}$.

In the following we consider, besides the Nash equilibrium $E^N(n)$, three other equilibria $E^{AK}(k,m)$, $E^{AM}(k,m)$ and $E^B(k,m)$, which result when either the countries in group K (AK-scenario) or in group M (AM-scenario) alone or the countries in both groups K and M (B-scenario) are forming a coalition and cooperatively determine their public good contributions within their group. For $L = K, M$ public good supply in $E^{AL}(k,m)$ will be denoted by $G^{AL}(k,m)$ and private consumption of a country in group K by $x_K^{AL}(k,m)$ and of a country in group M by $x_M^{AL}(k,m)$. In $E^B(k,m)$ the respective quantities are $G^B(k,m)$, $x_K^B(k,m)$ and $x_M^B(k,m)$.

As will be shown in the next sections, the equilibria $E^A(k,m)$ and $E^B(k,m)$ may be standalone allocations in which only one of the two groups K or M makes a positive contribution to the public good. Generally, for group $L = K, M$ of countries with size $l = k, m$ public good supply $G^S(l)$ in the standalone allocation $E^S(l)$ is characterized by the condition

$$G^S(l) + le(G^S(l), l) = lw. \tag{3}$$

Condition (3) follows from the maximization of utility $u(w - g_L, lg_L)$ of a member of L, i.e., under the assumption of equal burden-sharing among the countries in L when they cooperatively determine their public good contribution. The allocation $E^S(l)$ thus is the symmetric Lindahl solution for group L. In $E^S(l)$ private consumption of a member of group L is $x^S(l) = e(G^S(l), l)$ and $G^S(l) < \overline{G}_l$ holds (see Fig. 2).

In $E^S(l)$ a member of group L has utility $u^S(l) = u\left(x^S(l), G^S(l)\right)$, while utility of a free-rider F, which does not contribute to the public good and thus has private consumption $x_F^S(l) = w$, is $u_F^S(l) = u\left(w, G^S(l)\right)$.

Non-inferiority of the public good implies that $G^S(l)$ is increasing in group size l. In the Cobb-Douglas case with utility function $u(x_i, G) = x_i^\alpha G$ we especially have $G^S(l) = \frac{lw}{1+\alpha}$, $x^S(l) = \frac{\alpha w}{1+\alpha}$, $u^S(l) = l\alpha^\alpha\left(\frac{w}{1+\alpha}\right)^{1+\alpha}$ and $u_F^S(l) = \frac{lw^{1+\alpha}}{1+\alpha}$.

How the two equilibria based on one-sided and both-sided partial cooperation, respectively, look like and how they depend on group sizes k and m will now be explored in detail.

3 Unilateral Cooperation

In the AK-scenario group K jointly determines the public good contributions g_K^{AK} of each of its members collectively playing Nash against group M, whose member still choose their contributions g_M^{AK} non-cooperatively as in the standard Nash

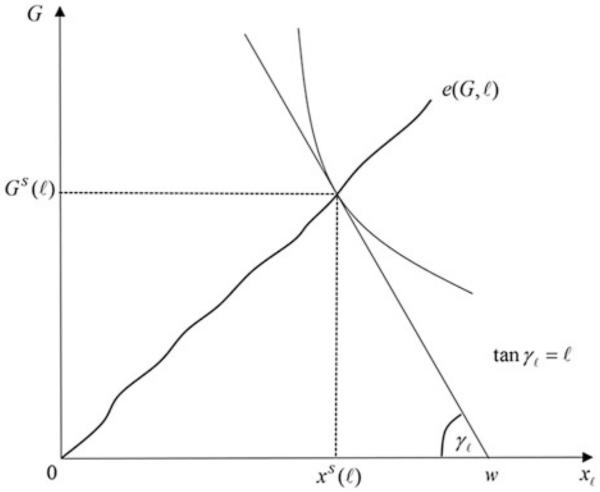


Fig. 2 Lindahl equilibrium positions

equilibrium. An interior Nash equilibrium $E^{AK}(k,m)$, in which both the countries in group K and group M are making positive contributions to the public good, can be characterized with the help of the AGA in a straightforward way. For that purpose let a public good supply level $G_I^{AK}(k,m)$ be given by the condition

$$G_I^{AK}(k,m) + ke\left(G_I^{AK}(k,m),k\right) + me\left(G_I^{AK}(k,m),1\right) = (k+m)w. \qquad (4)$$

Existence and uniqueness of $G_I^{AK}(k,m)$ follows from continuity and monotonicity of the expansion paths. Based on (4) we get a first characterization of $E^{AK}(k,m)$.

Proposition 1 If $E^{AK}(k,m)$ is interior we have $G^{AK}(k,m) = G_I^{AK}(k,m) < \overline{G}_1$, $x_K^{KA}(k,m) = e\left(G_I^{AK}(k,m),k\right) < w$ and $x_M^{AK}(k,m) = e\left(G_I^{AK}(k,m),1\right) < w$.

Proof On the one hand, if group K—again under the assumption of symmetric burden-sharing—determines its joint Nash reaction kg_K to the aggregate public good contributions mg_M of group M, it maximizes utility of each member $u\left(w - g_K, kg_K + mg_M\right)$. This in particular means that in case of cooperation every country in K faces $mrt = k$ as its individual marginal rate of transformation between the private and the public good. If the solution of this optimization leads to a positive public good contribution each country in K attains a position where $mrs = mrt = k$ holds, i.e., a position on the expansion path $e(G,k)$. On the other hand, the position of the countries in group M is still on the income expansion path $e(G,1)$ when their public good contribution is positive. In combination with the feasibility constraint (1) this shows that an interior $E^{AK}(k,m)$ is characterized by condition (4). Interiority implies that $e\left(G_I^{AK}(k,m),k\right) < e\left(G_I^{AK}(k,m),1\right) < w$ and thus $G_I^{AK}(k,m) < \overline{G}_1$.

QED

The equilibrium $E^{AK}(k,m)$, however, need not be interior. The solution may also be a corner equilibrium, in which either only the group K or the group M actively contributes to the public good. We first show that the latter case can be excluded, which is in contrast to the scenario in which the coalition acts as a Stackelberg leader (Buchholz et al. 2014).

Proposition 2 A corner equilibrium $E^{AK}(k,m)$, in which only the outsider group M contributes to the public good, can never occur.

Proof If only the members of group M made a positive contribution, public good supply would become $G^N(m)$, which in analogy to condition (2) is defined by the condition $G^N(m)+me\left(G^N(m),1\right)=mw$. Since $G^N(m)<\overline{G}_1<\overline{G}_k$, non-inferiority would imply that the *mrs* of the members of K at $(w,G^N(m))$ is smaller than k, so that—being confronted with $G^N(m)$—coalition K would have an incentive to contribute to the public good. This, however, is not compatible with an equilibrium outcome. QED

The other type of a corner solution is the symmetric standalone equilibrium $E^S(k)$ of coalition K. The following result provides criteria by which it can be determined whether the interior solution or the standalone allocation $E^S(k)$ emerges as the equilibrium outcome of the contribution game.

Proposition 3 $E^{AK}(k,m)$ is interior if $G_I^{AK}(k,m)<\overline{G}_1$ or, equivalently, if $G^S(k)<\overline{G}_1$. Otherwise, $E^{AK}(k,m)$ is the corner solution $E^S(k)$.

Proof

(i) As a first step we show that the two conditions for interiority are equivalent: Assume $G^S(k)<\overline{G}_1$. Since in this case $e\left(G^S(k),1\right)<w$, it follows from $G^S(k)+ke\left(G^S(k),k\right)=kw$ that

$$G^S(k)+ke\left(G^S(k),k\right)+me\left(G^S(k),1\right)<(k+m)w. \tag{5}$$

But as $G+ke(G,k)$ is increasing in G and $e(\overline{G}_1,1)=w$ we have

$$\overline{G}_1+ke\left(\overline{G}_1,k\right)+me(\overline{G}_1,1)>(k+m)w. \tag{6}$$

This implies that public good supply $G_I^{AK}(k,m)$, which satisfies (4), must lie in the interval $\left(G^S(k),\overline{G}_1\right)$ and thus, in particular, is smaller than $\overline{G}_1$. Conversely, if $G_I^{AK}(k,m)<\overline{G}_1$ it is shown by a similar argument that $G^S(k)<G_I^{AK}(k,m)$ holds.

(ii) Let $G_I^{AK}(k,m)<\overline{G}_1$. Then $e\left(G_I^{AK}(k,m),k\right)<w$ and $e\left(G_I^{AK}(k,m),1\right)<w$, and the countries in K and M are in an equilibrium position when they make the strictly positive public good contributions $g_{IK}^{AK}=w-e\left(G_I^{AK}(k,m),k\right)$ and $g_{IM}^{AK}=w-e\left(G_I^{AK}(k,m),1\right)$, respectively. This shows that the allocation as defined by condition (4) is a Nash equilibrium $E^{AK}(k,m)$. It is the only one

since the interior solution is unique, and group K's standalone allocation can be excluded: From part (i) of the proof we know that $G_I^{AK}(k,m) < \overline{G}_1$ implies $G^S(k) < \overline{G}_1$. Then the countries in M would have an incentive to contribute to the public good when being confronted with $G^S(k)$ so that coalition K's standalone allocation cannot be an equilibrium.

(iii) Assume that the conditions stated in the Proposition are not fulfilled, i.e., especially that $G^S(k) \geq \overline{G}_1$ holds. Given $G^S(k)$ the countries in group M then have no incentive to contribute to the public good since in their complete free-rider position $(w, G^S(k))$ their *mrs* is larger than their marginal rate of transformation $mrt = 1$. Therefore, K's standalone allocation $E^S(k)$ is a $E^{AK}(k,m)$, and it is the only one since an interior solution is excluded by Proposition 1 as $G_I^{AK}(k,m) \geq \overline{G}_1$ in this case. QED

It is a direct consequence of Proposition 3 that $E^{AK}(k,m)$ always exists and is unique. We now show how the fulfillment of the criteria described by Proposition 3 depends on the size of the two groups K and M.

Proposition 4 There is a coalition size $\overline{k}_A \geq 2$ so that, independent of the size of the outsider group m, $E^{AK}(k,m)$ is coalition K's standalone allocation $E^S(k)$ if and only if $k \geq \overline{k}_A$.

Proof Let $\overline{k}_A = \min_{k \geq 2} \left\{k \in \mathbb{N} : G^S(k) \geq \overline{G}_1\right\}$. Existence of $\overline{k}_A$ is ensured as there clearly exists some k for which the budget line $G = -kx_i + kw$ cuts the indifference curve passing through $\left(w, \overline{G}_1\right)$ so that $G^S(k) > \overline{G}_1$ definitely holds for such k.

The threshold level $\overline{k}_A$ defined in this way is independent of m as $\overline{G}_1$ is independent of m. Following Proposition 3, $\overline{k}_A$ has the required properties. Neglecting that $\overline{k}_A$ has to be a natural number it can be determined as in Fig. 3. QED

When all agents have the Cobb-Douglas utility function $u(x_i, G) = x_i^{\alpha} G$ we have $e(G,\rho) = \frac{\alpha}{\rho} G$ for some given $\rho > 0$ so that $\overline{G}_1 = \frac{w}{\alpha}$. As $G^S(k) = \frac{kw}{1+\alpha}$ the construction in the proof of Proposition 4 yields $\overline{k}_A = \min_{k \geq 2} \left\{k \in \mathbb{N} : k \geq \frac{1+\alpha}{\alpha}\right\}$. For $\alpha \geq 1$, we thus get $\overline{k}_A = 2$ so that an interior equilibrium never occurs in this case. But if $\alpha < 1$ is sufficiently close to zero interior Nash equilibria may emerge for arbitrarily large coalitions since $\lim_{\alpha \to 0} \frac{1+\alpha}{\alpha} = \infty$.

The equilibrium $E^{AM}(k,m)$, which results when only group M cooperates, is obtained from the considerations above by interchanging the roles between K and M. As seen by Proposition 4, it completely depends on the size of group M, which type of equilibrium arises in this case: If $m < \overline{m}_A := \overline{k}_A$ there is an interior solution, in which (omitting the argument (k, m) here) public good supply G^{AM} is G_I^{AM} given by the condition

$$G_I^{AM} + ke\left(G_I^{AM}, 1\right) + me\left(G_I^{AM}, m\right) = (k+m)w. \tag{7}$$

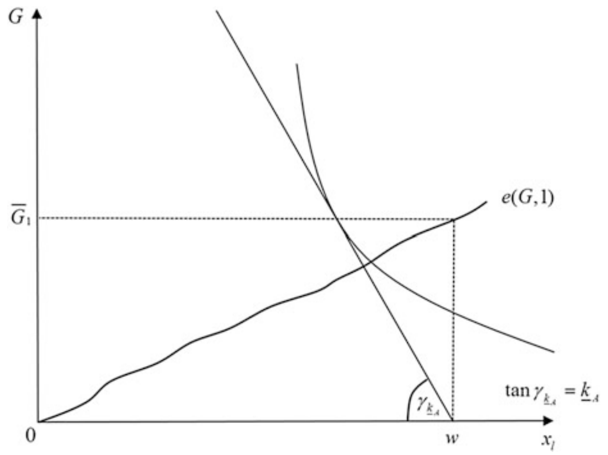


Fig. 3 Threshold $\overline{k}_A$

If, however, $m \geq \overline{m}_A$ the equilibrium outcome $E^{AM}(k,m)$ is the standalone equilibrium of group M, which implies $G^{AM} = G^S(m)$. The positions attained by the members of group K clearly are $\left(x_K^{AM}, G^{AM}\right) = \left(e\left(G_I^{AM}, 1\right), G_I^{AM}\right)$ if $m < \overline{m}_A$ and $\left(x_K^{AM}, G^{AM}\right) = \left(w, G^S(m)\right)$ if $m \geq \overline{m}_A$.

4 Bilateral Cooperation

In the second scenario both groups K and M internally cooperate when determining public good contributions of their members g_K^B and g_M^B, but play non-cooperatively against each other. Again, the AGA can be applied to characterize the interior equilibria of this game with two-sided partial cooperation. For that purpose let an allocation $E_I^B(k,m)$ with public good supply $G_I^B(k,m)$ be given by the equation

$$G_I^B(k,m) + ke\left(G_I^B(k,m),k\right) + me\left(G_I^B(k,m),m\right) = (k+m)w. \tag{8}$$

Existence and uniqueness of $G_I^B(k,m)$ again follows from the properties of expansion paths. In complete analogy to Proposition 1 we obtain the following result:

Proposition 5 If $E^B(k,m)$ is interior we have $G^B(k,m) = G_I^B(k,m) < \min\left\{\overline{G}_k, \overline{G}_m\right\}$, $x_K^B(k,m) = e\left(G_I^B(k,m),k\right) < w$ and $x_M^B(k,m) = e\left(G_I^{AK}(k,m),m\right) < w$.

Utilities are of the members of group K and the members of group M then are denoted by $u_I^{BK}(k,m)$ and $u_I^{BM}(k,m)$, respectively. In contrast to the case of unilateral cooperation it now becomes possible that the outcome is the standalone equilibrium of each group either.

Proposition 6 $E^B(k,m)$ is interior if $G_I^B(k,m) < \min\left\{\overline{G}_k, \overline{G}_m\right\}$ or, equivalently, if $G^S(k) < \overline{G}_m$ and $G^S(m) < \overline{G}_k$. If, instead, $G^S(m) \geq \overline{G}_k$, $E^B(k,m)$ is the corner solution $E^S(m)$, and if $G^S(k) \geq \overline{G}_m$, $E^B(k,m)$ is the corner solution $E^S(k)$.

Proof

(i) It is shown in an analogous way as in the proof of Proposition 3 that the two conditions for interiority are equivalent and that $E^B(k,m) = E_I^B(k,m)$ in this case.
(ii) Let $G^S(k) \geq \overline{G}_m$. Given $G^S(k)$ the countries in M then have no incentive to make a collective contribution to the public good since at $(w, G^S(k))$ their *mrs* is larger than their $mrt = m$, so that $E^S(k)$ is an equilibrium. It is the only one since, on the one hand, an interior solution is excluded by (i). On the other hand, the standalone allocation of group M is not possible as $G^S(m) < \overline{G}_m \leq G^S(k) < \overline{G}_k$. But $G^S(m) < \overline{G}_k$ implies that the countries in K would have an incentive to contribute when they are confronted with public good supply $G^S(m)$. The case $G^S(m) \geq \overline{G}_k$ is treated in the same way. QED

Existence of and uniqueness of $E^B(k,m)$ follows from Proposition 6 as its three cases cover all possible situation and are mutually exclusive since $G^S(k) \geq \overline{G}_m$ implies $G^S(m) < \overline{G}_k$. We now show how the fulfillment of the criteria described by Proposition 6 depends on the size of the two groups K and M.

Proposition 7 For any $m \geq 2$ there exist threshold levels $\underline{k}_B(m)$ and $\overline{k}_B(m)$, for which $\overline{k}_B(m) > m > \underline{k}_B(m) \geq 0$ holds and which are both increasing in m, so that

(i) $E^B(k,m) = E^S(m)$ if $k \leq \underline{k}_B(m)$.
(ii) $E^B(k,m) = E_I^B(k,m)$ if $\underline{k}_B(m) < k < \overline{k}_B(m)$.
(iii) $E^B(k,m) = E^S(k)$ if $k \geq \overline{k}_B(m)$.

Furthermore, $\lim_{m\to\infty} \underline{k}_B(m) = \lim_{m\to O} \overline{k}_B(m) = \infty$.

Proof Let $\underline{k}_B(m) = \max\left\{k \in \mathbb{N} : \overline{G}_k \leq G^S(m)\right\}$ and $\overline{k}_B(m) = \min\left\{k \in \mathbb{N} : G^S(k) \geq \overline{G}_m\right\}$. The assertions in (i), (ii) and (iii) then are a direct consequence of Proposition 6. Monotonicity of $\underline{k}_B(m)$ and $\overline{k}_B(m)$ follows as $\overline{G}_m$ and $G^S(m)$ are monotone increasing in m. That any given m lies in the interiority range follows since

$$\overline{G}_m + 2e\left(\overline{G}_m, m\right) = \overline{G}_m + 2mw > 2mw \tag{9}$$

so that for $k = m$ the Eq. (8) has a solution $G_I^B(m,m) < \overline{G}_m$.

To show $\lim_{m\to\infty} \underline{k}_B(m) = \infty$ we prove that $G^S(m)$ cannot be bounded above. Otherwise, if there were an upper bound $\overline{G}^S < \infty$ of $G^S(m)$, we would get a contradiction by choosing some m for which the budget line with slope m starting at $(w,0)$ cuts the indifference curve passing through $\left(w, \overline{G}^S\right)$. QED

The set of k, for which according to (i) a standalone equilibrium of group M results, may be empty. The sets of k as defined by (ii) and (iii) in Proposition 7, however, are non-empty for all $m \geq 2$. For (ii) this follows by letting $k = m$ and for (iii) through an argument as in the proof of Proposition 4.

In the Cobb-Douglas case with the utility function $u(x_i, G) = x_i^{\alpha} G$ we especially have $\overline{G}_k = \frac{kw}{\alpha}$, $\overline{G}_m = \frac{mw}{\alpha}$, $G^S(k) = \frac{kw}{1+\alpha}$ and $G^S(m) = \frac{mw}{1+\alpha}$. Real-numbered threshold levels $\underline{k}_B(m)$ and $\overline{k}_B(m)$ then are defined by $\overline{G}_{\underline{k}_B(m)} = \frac{\underline{k}_B(m)w}{\alpha} = \frac{mw}{1+\alpha} = G^S(m)$ and $G^S\left(\overline{k}_B(m)\right) = \frac{\overline{k}_B(m)w}{1+\alpha} = \frac{mw}{\alpha} = \overline{G}_m$, respectively, which gives

$$\underline{k}_B(m) = \frac{\alpha}{1+\alpha} m \quad \text{and} \quad \overline{k}_B(m) = \frac{1+\alpha}{\alpha} m. \tag{10}$$

The both equations in (10) also show that—given some α—an increase of group size m enlarges the sets of k, leading to a standalone equilibrium of group M and an interior equilibrium, while it reduces the set of k, leading to a standalone equilibrium of group K. Moreover, an increasing preference for the private good α makes both types of standalone equilibria more likely and an interior solution more unlikely.

If $\underline{k}_B(m) < k < \overline{k}_B(m)$ a straightforward application of condition (8) to the Cobb-Douglas case gives $G_I^B(k,m) = \frac{(k+m)w}{1+2\alpha}$ for public good supply and $x_I^{BK}(k,m) = e\left(G_I^B, k\right) = \frac{\alpha(k+m)w}{k(1+2\alpha)}$ and $x_I^{BM}(k,m) = e\left(G_I^B, m\right) = \frac{\alpha(k+m)w}{m(1+2\alpha)}$ for private consumption of the countries in K and M, respectively.

5 Incentives for Coalition Formation

We now explore whether in a two-stage game the group K has an incentive to form a coalition at stage 1 given that the members of group M either have formed a coalition or not. Therefore we have to compare utility of the countries in K between the allocations $E^N(k,m)$ and $E^{AK}(k,m)$ on the one hand and between allocations $E^B(k,m)$ and $E^{AM}(k,m)$ on the other. Albeit for different reasons in both cases group K will have a weaker (stronger) incentive for partial cooperation when it is small (large) while group M is large (small).

5.1 The Optimal Coalition Formation Decision of Group K When Group M Does not Cooperate

According to Proposition 2 we have to distinguish between the cases, in which $E^{AK}(k,m)$ is an interior solution or $E^{AK}(k,m)$ is the standalone allocation of group K, i.e., if according to Proposition 4 either $k < \overline{k}_A$ or $k \geq \overline{k}_A$.

For the treatment of the first case $k < \overline{k}_A$ let for any $\xi \in [0, k]$ public good supply $\tilde{G}_I^{AK}(\xi, m)$ be given by

$$\tilde{G}_I^{AK}(\xi, m) + ke\left(\tilde{G}_I^{AK}(\xi, m), \xi\right) + me\left(\tilde{G}_I^{AK}(\xi, m), 1\right) = (k+m)w. \tag{11}$$

We obtain $G^N(k+m) = \tilde{G}_I^{AK}(1, m)$ and $G_I^{AK}(k, m) = \tilde{G}_I^{AK}(k, m)$ as special cases. The function $\tilde{u}_I^{AK}(\xi)$ indicates utility a member of K attains at some $\xi \in [1, k]$, where its private consumption is $\tilde{x}_I^{AK}(\xi, m) = e\left(\tilde{G}_I^{AK}(\xi, m), \xi\right)$. Omitting arguments, using e_1^K and e_2^K as abbreviations for the two partial derivatives of the expansion path $e(G, \rho)$ at $G = \tilde{G}_I^{AK}(\xi, m)$ and $\rho = \xi$ and e_1^M for the first partial derivative of the expansion path $e(G, 1)$ at $G = \tilde{G}_I^{AK}(\xi, 1)$, and observing that $\xi = \frac{u_1\left(x_I^{AK}, G_I^{AK}\right)}{u_2\left(x_I^{AK}, G_I^{AK}\right)}$ we get, similarly as in Boadway and Hayashi (1999, p. 629),

$$\frac{\partial \tilde{u}_I^{AK}}{\partial \xi} = u_2\left(\left(\xi e_1^K + 1\right)\frac{\partial \tilde{G}_I^{AK}}{\partial \xi} + \xi e_2^K\right). \tag{12}$$

Total differentiation of (11) gives

$$\frac{\partial \tilde{G}_I^{AK}}{\partial \xi} = \frac{-ke_2^K}{1 + ke_1^K + me_1^M}. \tag{13}$$

Combining (12) and (13) and having in mind that the normality assumption implies $e_1^K > 0$, $e_1^M > 0$ (so that the denominator of (13) is positive) and $e_2^K < 0$ yields

$$\frac{\partial \tilde{u}_I^{AK}}{\partial \xi} < 0 \quad \text{if and only if} \quad k < \xi\left(1 + me_1^M\right) > 0. \tag{14}$$

If the slope of the expansion path w.r.t. G is bounded from below by $\underline{e}_1$ condition (14) leads to the following result:

Proposition 8 If group M does not cooperate, group K does not benefit from forming a coalition if

$$k < \min\left\{\overline{k}_A, 1 + m\underline{e}_1\right\}. \tag{15}$$

Proof Condition (15) implies that (14) is fulfilled for all $\xi \in [0, k]$ so that $\tilde{u}_I^{AK}(\xi)$ is decreasing for all $\xi \in [1, k]$, which yields $u_K^{AK}(k+m) = \tilde{u}_I^{AK}(k) < \tilde{u}_I^{AK}(1) = u^N(k+m)$. QED

For the treatment of the second case where $k \geq \overline{k}_A$ we first define

$$k_A^* = \min_{k \geq 2}\left\{k \in \mathbb{N} : u^S(k) \geq u\left(w, \overline{G}_1\right)\right\}. \tag{16}$$

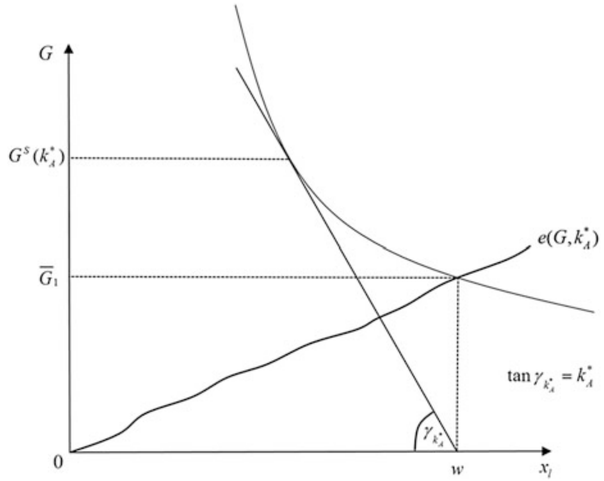


Fig. 4 Threshold k_A^*

In a x_i–G-diagram the threshold level k_A^* is obtained as the minimum value of group size k for which the budget line with slope k cuts the indifference curve passing through the point $\left(w, \overline{G}_1\right)$. Neglecting again that k_A^* has to be an integer, it is characterized by the tangent to this indifference curve passing through the point $(w, 0)$ (see Fig. 4). Clearly, k_A^* exists and $k_A^* > \overline{k}_A$.

For any integer $m \geq 2$ we furthermore define

$$k_A^*(m) = \min_{k \geq 2} \left\{k \in \mathbb{N} : u^S(k) > u^N(k+m)\right\}. \tag{17}$$

Since $x^N(k+m) < w$ and $G^N(k+m) < \overline{G}_1$ and thus $u^S\left(k_A^*\right) > u^N(k+m)$ for all k and m, $k_A^*(m)$ exists and $k_A^*(m) \leq k_A^*$ holds for all $m \geq 2$. Based on (17) conditions can be provided, which ensure that coalition formation is the optimal reaction of group K.

Proposition 9 If group M does not cooperate, the members of group K will benefit from forming a coalition if

$$k \geq \widehat{k}_A(m) := \max\left\{\overline{k}_A, k_A^*(m)\right\}. \tag{18}$$

For any $k \geq k_A^*$ holds coalition formation is profitable for group K even independent of the size m of group M.

Proof The result directly follows from the definitions of $\overline{k}_A$, $k_A^*(m)$ and k_A^*. QED

Both Propositions 8 and 9 confirm that coalition building is fostered, when group K is large and the outsider group M is small.

5.2 The Optimal Coalition Formation Decision of Group K When Group M Cooperates

To explore whether group K has an incentive to cooperate, when group M does we have to compare utility of its members in the allocations $E^{AM}(k,m)$ and $E^{B}(k,m)$. To simplify the exposition we concentrate on the situation where the standalone solution of group M emerges as Nash equilibrium at the contribution stage when group K does not cooperate, i.e., $E^{AM}(k,m) = E^{S}(m)$ and $m \geq \overline{m}_A$, and the members of K have utility $u_F^S(m)$ when they do not cooperate. First of all, it is straightforward to provide a condition which ensures that group K prefers non-cooperation in this case.

Proposition 10 If group M cooperates and $m \geq \overline{m}_A$ group K does not benefit from forming a coalition if $k < \underline{k}_B(m)$.

Proof It follows from Proposition 7 that a cooperating group K does not want to make a positive contribution to the public good if $k < \underline{k}_B(m)$. Thus $E^{B}(k,m) = E^{AM}(k,m)$ so that coalition formation would not make a change. QED

When $\underline{k}_B(m) < k < \overline{k}_B(m)$ and hence $E^{B}(k,m)$ is interior, cooperation K may also not pay for the members of group K. So one can expect that cooperation will reduce utility of the countries in K below their utility in the standalone equilibrium of group M if group size k lies only slightly above the lower interiority threshold $\underline{k}_B(m)$. To show this let for a continuous variable $\xi \geq \underline{k}_B(m)$ public good supply $G_I^B(\xi,m)$ be defined by

$$G_I^B(\xi,m) + \xi e\left(G_I^B(\xi,m),\xi\right) + me\left(G_I^B(\xi,m),m\right) = (\xi+m)\,w. \tag{19}$$

If $u_I^{BK}(\xi,m)$ denotes utility, which a member of coalition K then attains at ξ, we get the following result where $\underline{\xi}_B(m)$ is given by $G_{\underline{\xi}_B}(m) = G^s(m)$.

Proposition 11 There is some threshold $\underline{\chi}$ so that for all $\xi \in \left[\underline{\xi}_B(m),\underline{\chi}\right]$ utility $u_I^{BK}(\xi,m)$ is falling in ξ.

Proof See the Appendix. QED

A straightforward implication of Proposition 11 is that under some additional conditions cooperation of group K will not pay for its members whenever the equilibrium at the contribution stage is interior. To show this first define for any $m \geq 2$

$$k_B^*(m) := \min_{k\geq 2}\left\{k \in \mathbb{N} : u^S(k) \geq u_F^S(m)\right\}. \tag{20}$$

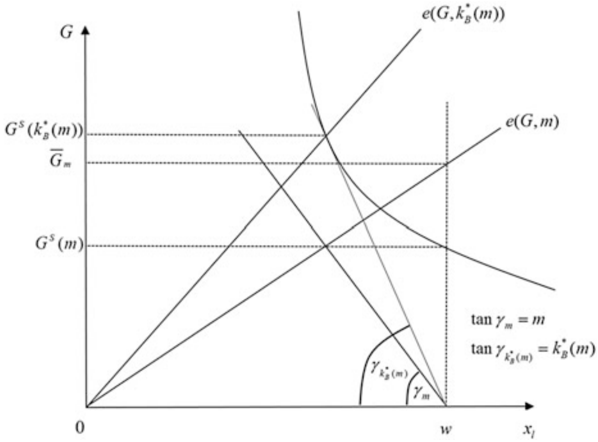


Fig. 5 Threshold k_B^*

This definition means that for all $k \geq k_B^*(m)$ utility of a country in K is higher in K's standalone solution than it would be as a free-rider in M's standalone solution. Since $u^S(m) < u_F^S(m)$ and $u^S(k)$ is increasing in k we have $k_B^*(m) > m$ (see Fig. 5). As $u_F^S(m)$ is increasing in m, $k_B^*(m)$ is non-decreasing. We then have the following result:

Proposition 12 Assume $m \geq \overline{m}_A$, that $u_I^{BK}(\xi, m)$ is a convex function of ξ and that $k_B^*(m) > \overline{k}_B(m)$. For all k with $\underline{k}_B(m) < k < \overline{k}_B(m)$ then group K does not benefit from forming a coalition when group M cooperates.

Proof As $u^S(k)$ is increasing in k and $k_B^*(m) > \overline{k}_B(m)$ is assumed it follows that $u_I^{BK}\left(\overline{\xi}_B(m), m\right) = u^S\left(\overline{\xi}_B(m)\right) < u^S\left(\overline{k}_B(m)\right) \leq u_F^S(m)$ where $\overline{\xi}_B(m)$ is defined by $\overline{G}_{\overline{\xi}_B(m)} = \overline{G}_m$. According to Proposition 11 utility $u_I^{BK}(\xi, m)$, which a country in K attains in an interior solution, is falling on an interval $\left[\underline{k}_B(m), \underline{\chi}\right]$. As the convexity assumption implies that the function $u_I^{BK}(\xi, m)$ can at most have one minimum in $\left[\underline{k}_B(m), \overline{k}_B(m)\right]$ and since $u_I^{BK}\left(\underline{k}_B(m), m\right) = u_F^S(m)$, we have $u_I^{BK}(k, m) < u_F^S(m)$ for all k with $\underline{k}_B(m) < k < \overline{k}_B(m)$. QED

Taken together Propositions 10 and 12 imply, that in the situation underlying Proposition 12, non-cooperation is group K's best reply to cooperation of group M for all $k \leq \overline{k}_B(m)$. As in the case, where group M does not cooperate, coalition building, however, becomes profitable if group K is sufficiently large.

Based on $k_B^*(m)$ as defined by (20) we can provide a sufficient condition for coalition building by group K when M has formed a coalition.

Proposition 13 If group M cooperates group K benefits from forming a coalition if

$$k > \widehat{k}_B(m) := \max\left\{\overline{k}_B(m), k_B^*(m)\right\} > m. \tag{21}$$

$\widehat{k}_B(m)$ is non-decreasing in m and $\lim_{m\to\infty} \widehat{k}_B(m) = \infty$.

Proof The assertions directly follow from the definitions and the monotonicity of $\overline{k}_B(m)$ and $k_B^*(m)$. That $\lim_{m\to\infty} \widehat{k}_B(m) = \infty$ is a consequence of Proposition 7, i.e., of $\lim_{m\to\infty} \overline{k}_B(m) = \infty$. QED

Proposition 13 shows that the incentives for group K to form a coalition are larger if the group M is small. However, it has to be emphasized that—unlike the case where group M does not cooperate—it is not possible that a group K of some given size k always wants to cooperate irrespective of the size of M.

Proposition 14 For any $k \geq 2$ there exists a $\tilde{m}(k)$ so that a group K of size k does not benefit from forming a coalition if the size of the cooperating group M is $m \geq \tilde{m}(k)$.

Proof The assertion again directly follows from Proposition 10 since $\lim_{m\to\infty} \underline{k}_B(m) = \infty$ as stated in Proposition 7. QED

5.3 A Comparison of the Optimal Reactions

The results of the previous sub-sections show that, concerning group K's coalition formation decision, some similarities but also some substantial differences between the two cases considered above exist.

On the one hand, there is a common tendency that cooperation is not profitable for group K if it is small and that it becomes always profitable if it is big enough—irrespective of whether the other group M acts as a coalition or not. But on the other hand, coalition formation of group K is much more likely if the other group M does not cooperate, which is reflected by some of the results formulated above: If the members of M choose their public good contributions non-cooperatively any group K whose size exceeds some minimum level wants to form a coalition independent of how large group M is (see Proposition 9). When M, however, cooperates such a lower bound does not exist. Rather, for any k we can find a sufficiently large group m so that group K has no incentive to form a coalition (see Proposition 14). Moreover, the minimum coalition size beyond which cooperation is definitely in the interest of the countries in group K is higher when group M cooperates than when it does not. This result is stated by the following Proposition.

Proposition 15 Assume $m > \overline{m}_A$. For the threshold levels defined in (18) and (21), which provide sufficient conditions for coalition formation of group K, $\widehat{k}_B(m) > \widehat{k}_A(m)$ holds.

Proof On the one hand, $m \geq 2$ implies $\overline{G}_m > \overline{G}_1$ and thus $\overline{k}_B(m) > \overline{k}_A$. On the other hand, $m > \overline{m}_A$ gives $G^S(m) > \overline{G}_1$ so that $u\left(w, \overline{G}_1\right) < u\left(w, G^S(m)\right) = u_F^S(m)$. Comparing (16) and (20) then shows $k_B^*(m) > k_A^* \geq k_A^*(m)$. Taken together, we obtain $\widehat{k}_B(m) = \max\left\{\overline{k}_B(m), k_B^*(m)\right\} > \widehat{k}_A(m) = \max\left\{\overline{k}_A, k_A^*(m)\right\}$. QED

6 Nash Equilibria at the Coalition Formation Stage: Some General Conditions

Changing roles and taking the reactions of group M also into account now allows us to derive some general results concerning the type of the Nash equilibria that emerges at the first coalition formation stage of the entire two-stage game. So Proposition 9 has shown that, if the size of a group is large enough, it will always react by cooperating when the other group does not cooperate, i.e., $N_M \to C_K$ and $N_K \to C_M$. Concerning the reaction when the other group cooperates, our general results suggest that the members of a small group prefer to standalone when the other group is large, i.e., $C_M \to N_K$ or $C_K \to N_M$. To formulate conditions for a unique Nash equilibrium in this case consider the function $\widehat{m}_B(k)$, which is defined by (21) through interchanging k and m, and then its inverse $\breve{k}_B(m)$, which in an m–k-diagram is obtained by mirroring $\widehat{k}_B(m)$ on the 45^0-line. As $\widehat{m}_B(k) > k$ for all k we get $\breve{k}_B(m) < m$. Furthermore, let $\underset{\smile}{k}_B(m) := \min\left\{\underline{k}_B(m), \breve{k}_B(m)\right\} < m$.

Proposition 16 If $k, m > k_A^*$ and $k < \underset{\smile}{k}_B(m) < m$ there is a unique Nash equilibrium (N_K, C_M) at the coalition formation stage in which the larger group M is willing to form a coalition while the smaller group K is not. At the second stage group M's standalone equilibrium $E^S(m)$ results.

Proof As noted above we have $N_K \to C_M$ as $m \geq m_A^* = k_A^*$. But given $k < \underline{k}_B(m)$ Proposition 10 says that $C_M \to N_K$ so that (N_K, C_M) is a Nash equilibrium. It is the only one since it follows from Proposition 13 and the construction of $\underset{\smile}{k}_B(m)$ that for all (m, k) with $k < \underset{\smile}{k}_B(m)$ group M wants to cooperate when group K does, i.e., $C_K \to C_M$. Under the conditions underlying the Proposition group K clearly must be smaller than group M as $\underline{k}_B(m) < m$. The assertion concerning the second stage follows from Proposition 4 since $k_A^* > \overline{k}_A$. QED

To put it in another way: In the situation as given in Proposition 16 the payoff structures of both groups are different: While group K has payoffs as in a chicken game those of group M are as in a harmony game. This constellation gives a unique Nash equilibrium at the coalition formation stage.

If, in contrast to the situation described in Proposition 16, the sizes of both groups do not diverge too much, each group clearly will show the same reaction when the other group has built a coalition, i.e., it will either choose non-cooperation or cooperation. Assume again that groups are large enough so that non-cooperation of one group is responded by cooperation of the other. If the best reaction to cooperation is non-cooperation then a *chicken game* with two asymmetric Nash equilibria emerges. Under the same assumptions as made in Proposition 12 such an outcome may result when the Nash equilibrium at second stage is interior.

Proposition 17 Let $k, m > k_A^*$ and assume that $u_I^{BK}(\xi, m)$ is a convex function of ξ and that $k_B^*(m) > \overline{k}_B(m)$. Then for all (k, m) with $\underline{k}_B(m) < k < \overline{k}_B(m)$ there are two Nash equilibria (C_K, N_M) and (N_K, C_M). At the second stage then either the standalone equilibrium $E^S(k)$ of group K or the standalone equilibrium $E^S(m)$ of group M result.

Proof Concerning the first stage, $k, m > k_A^*$ yields the reactions $N_K \rightarrow C_M$ and $N_M \rightarrow C_K$. From Proposition 12 we have $C_M \rightarrow N_K$ for all k with $\underline{k}_B(m) < k < \overline{k}_B(m)$. By reversing roles and observing symmetry we also obtain $C_K \rightarrow N_M$. Concerning the second stage, the assertion again follows from Proposition 4. QED

As always in the case of a chicken game a unique solution emerges when we assume that one of the two groups, say group M, either commits to cooperation or to non-cooperation. Regarding public good supply we then have the following result:

Proposition 18 Under the same assumptions as in Proposition 17 and given any combination (k, m) with $m < k < \overline{k}_B(m)$ public good supply will be higher when group M commits to non-cooperation than when it commits to cooperation.

Proof It follows from Proposition 17 that (N_K, C_M) with public good supply $G^S(m)$ results when group M commits to cooperation. When group M instead commits to non-cooperation group K forms a coalition and public good supply becomes $G^S(k)$. Since $k < m$ we then have $G^S(k) < G^S(m)$. QED

Proposition 18 shows that a cooperation of a relatively small coalition may motivate the members of a much larger group to choose the standalone strategy and to become free-riders. In this case the presence of the smaller group M leads to a public good supply which is lower than in the situation in which M were absent. Non-cooperation of the smaller group M instead creates an incentive for the larger group K to build a coalition and then to provide the public good at a higher level, so that in this sense less goodwill by group M is advantageous for public good provision. The difference between the public good supply levels in the two cases considered in Proposition 18 may even become quite large which is shown through an example in the subsequent section.

The result stated in Proposition 18 resembles some well-known paradoxical effects that may occur in the context of climate policy: Leading behavior by a group of countries aiming at improving global environmental quality [e.g., through unilateral increases of abatement efforts as in Hoel (1991), or to carbon taxes with a rapid increase of tax rates as in Sinn (2012)] can have the counterintuitive effect that in the end environmental quality deteriorates. In this paper, we have seen that such an adverse outcome may also arise from coalition formation decisions by groups of countries.

Looking at the sequential version of the chicken game [see also Foucart and Wan (2016)], the outcome as usual depends on which of the two groups is the first mover.

Proposition 19 Under the assumptions of Propositions 17 and 18 the outcome always is unilateral cooperation (C_K, N_M) with public good supply $G^S(k)$ when the smaller group M moves first. When the larger group K is the first mover, unilateral cooperation (N_K, C_M) with public good supply $G^S(m)$ results.

Proof If group M is moving first it follows from Proposition 17 that $E^S(m)$ results at the second stage when it forms a coalition and $E^S(k)$ when it does not. In the first case utility of a member of M is $u^S(m)$ and in the second case it is $u^S_F(k)$. As $m < k$ and thus $G^S(m) < G^S(k)$ we have $u^S(m) = u\left(w - \frac{G^S(m)}{m}, G^S(m)\right) < u(k, G^S(k)) = u^S_F(k)$, so that group M is better off by not cooperating. The second part of the assertion follows from Definition (20) since $k < k^*_B(m)$ implies $u^S_F(m) > u^S(k)$. QED

As $G^S(m) < G^S(k)$ then, in the situation underlying Proposition 19, equilibrium public good supply is smaller when the larger group K moves first. This result is surprising, since the intuition might suggest the opposite outcome.

The assumptions made in Proposition 17, which lead to a chicken game, do not hold in any case. Hence, it is also possible that with fairly equal group sizes the best reply to cooperation becomes cooperation, which leads to a *harmony game* with the unique Nash equilibrium (C_K, C_M). But in other rather special cases a *prisoner dilemma* game with the unique Nash equilibrium (N_K, N_M) or an *assurance game* with the two Nash equilibria (N_K, N_M) or (C_K, C_M) may result too. In the next section, where Cobb-Douglas preferences are assumed, we present examples for all these possible outcomes.

7 Nash Equilibria at the Coalition Formation Stage: The Cobb-Douglas Example

Let again each country's utility function be $u(x_i, G) = x_i^\alpha G$. To simplify the exposition, we normalize the initial endowment of each country to $w = 1$. We also no longer make a distinction between natural numbers k and the continuous variable ξ. The details of the partly tedious calculations, which underlie our results in this specific case, can be obtained from the authors upon request.

Firstly, we assume $\alpha \geq 1$ so that $\overline{k}_A = \frac{1+\alpha}{\alpha} \leq 2$. According to Proposition 4 then for any $k \geq 2$ unilateral cooperation of group K leads to this group's standalone solution $E^S(k)$ at the contribution stage, where utility of a country in K is

$$u^S(k) = \frac{k\alpha^\alpha}{(1+\alpha)^{1+\alpha}}. \tag{22}$$

When both groups do not cooperate each country's utility in the standard Nash equilibrium is

$$u^N(k+m) = \alpha^\alpha \left(\frac{k+m}{1+\alpha(k+m)} \right)^{1+\alpha}. \tag{23}$$

To determine $k_A^*(m)$ and thus the optimal cooperation decision of group K we would have to compare the utility levels given by (22) and (23), which however, does not allow for a closed-form solution. But it is possible to give an explicit solution for the threshold level k_A^* which is defined by the condition

$$u^S\left(k_A^*\right) = \frac{k_A^* \alpha^\alpha}{(1+\alpha)^{1+\alpha}} = \frac{1}{\alpha} = u\left(w, \overline{G}_1\right). \tag{24}$$

Thus

$$k_A^* = \left(\frac{1+\alpha}{\alpha} \right)^{1+\alpha}. \tag{25}$$

k_A^* is decreasing in α and converges to the Euler number e when α goes to infinity. If $\alpha \geq 1$ we thus have $k_A^* \leq 4$ so that in this case for all $k \geq 4$ coalition formation is the best response of group K when the other group M does not cooperate. In the following we will concentrate on this case. If the other group M instead has formed a coalition Proposition 10 shows that group K's best reaction is non-cooperation if $k \leq \underline{k}_B(m) = \frac{\alpha}{1+\alpha} m$.

To infer group K's coalition formation decision when $\underline{k}_B(m) < k < \overline{k}_B(m)$ and an interior equilibrium emerges at the second stage we have to consider

$$u_I^{BK}(k,m) = \left(\frac{\alpha}{k}\right)^{\alpha}\left(\frac{k+m}{1+2\alpha}\right)^{1+\alpha}, \tag{26}$$

which is convex in k as further calculations show. From

$$u^S\big(k_B^*(m)\big) = \frac{k_B^*(m)}{\alpha}\left(\frac{\alpha}{1+\alpha}\right)^{1+\alpha} = \frac{m}{1+\alpha} = u_F^S(m). \tag{27}$$

we get

$$k_B^*(m) = \left(\frac{1+\alpha}{\alpha}\right)^{\alpha} m, \tag{28}$$

so that for $\alpha > 1$ we have $k_B^*(m) > \overline{k}_B(m) = \frac{1+\alpha}{\alpha}m$. Proposition 12 thus provides that then K's optimal reaction to cooperation of M is non-cooperation also when $\underline{k}_B(m) < k < \overline{k}_B(m)$. For $\alpha = 1$ direct calculations provide the same result. Non-cooperation also is the best response of K to cooperation of group M if $k > \overline{k}_B(m)$ but $k < k_B^*(m)$ while group K will choose cooperation as soon as $k \geq k_B^*(m)$. Note that the set $\left\{k \in \mathbb{N} : \overline{k}_B(m) < k < k_B^*(m)\right\}$ may be empty, especially if α is not far above 1 and m is small.

Based on these results we now provide a complete description of the Nash equilibria at the coalition formation stage for the special case $\alpha = 1$. We then have $\overline{k}_A = 2$, $k_A^* = 4$, $\underline{k}_B(m) = \frac{1}{2}m$ and $\overline{k}_B(m) = k_B^*(m) = 2m$, which allows us to determine the Nash equilibria for the utmost part of all combinations of group sizes k and m: If $k, m \geq 4$ and $k < \frac{1}{2}m$ it follows from Proposition 16 that (N_K, C_M) is the unique Nash equilibrium at the coalition formation stage since $\widehat{m}_B(k) = \overline{m}_B(k) = m_B^*(k) = 2k$ and thus $\underset{\smile}{k}_B(m) = \frac{1}{2}m = \underline{k}_B(m)$. Analogously, (C_K, N_M) is the unique Nash equilibrium if $k > 2m$. If, however, $\frac{1}{2}m \leq k \leq 2m$ holds, Proposition 17 shows that a chicken game with the two Nash equilibria (N_K, C_M) and (C_K, N_M) emerges at the coalition formation stage.

The case $k = 2, 3$ has to be treated separately: If then $m \geq 4$ the unique Nash equilibrium at the coalition formation stage is (N_K, C_M) and $m = 2, 3$ and if $k \geq 4$ it is (C_K, N_M). We thus obtain the partitioning of the k-m-space as depicted in Fig. 6.

For the remaining combinations (2,2), (2,3), (3,2) and (3,3) inside the box of Fig. 6 we infer that for (2,3) the unique Nash equilibrium is (N_K, C_M) and for (3,2) it is (C_K, N_M). At (3,3) there are the two Nash equilibria (N_K, C_M) and (C_K, N_M). In the case (2,2) it is the dominant strategy of each group not to build a coalition so that there is a prisoners' dilemma at the coalition formation stage with (N_K, N_M) as the only Nash equilibrium.

For an illustration of Proposition 18 let $k = 2m-1$. Then $k < \overline{k}_B(m) = 2m$ so that equilibrium public good supply is $G^S(m) = \frac{m}{2}$ when the smaller group M commits

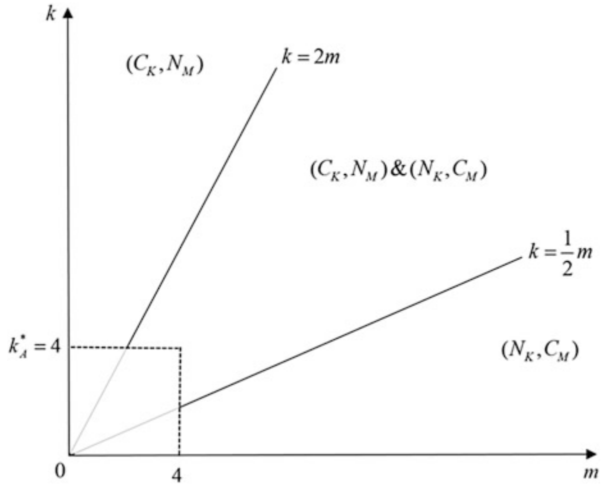


Fig. 6 The regions of subgame-perfect equilibria

to cooperation. But public good supply would become $G^S\,(2m - 1) = m - \frac{1}{2}$ when group M commits to non-cooperation. Hence, if group M forms a coalition this almost halves public good supply as compared to the outcome where it is not willing or unable to do so. If in the same situation coalition formation is a sequential game, in which the larger group K moves first, then also (N_K, C_M) with the lower public good supply $G^S(m)$ results. This illustrates Proposition 19.

For other Cobb-Douglas preferences also other game types at the coalition formation stage may occur. Let, e.g., $\alpha = \frac{1}{2}$ and $k = m \geq 3$. Then (C_K, C_M) is a Nash equilibrium at the first coalition-formation stage. If, in addition, $k = m > 6$ coalition-building becomes the dominant strategy for both countries so that there is a harmony game, in which bilateral cooperation (C_K, C_M) even is the unique Nash equilibrium. However, if $k = m = 3$ an assurance game results, which has the two symmetric Nash equilibria (N_K, N_M) and (C_K, C_M).

8 Conclusions

The results derived in this paper pour some cold water on the optimistic expectation, which is connected with the bottom-up-approach in climate policy. Rather, partial cooperation by one group of countries may make it less likely that the members of the other group are willing to build their own coalition, particularly if the cooperating coalition is large. But, more surprisingly, even relatively large groups might prefer to become free-riders in the standalone allocation brought about by a much smaller group. This causes the danger that the level of public good supply in the subgame-perfect equilibrium is lower if a smaller group is present than it would be if the larger group were alone. This undesired effect on the level of global public good provision occurs in many of our scenarios if the smaller group commits to coalition formation or if in a leader-follower version of the coalition formation game the larger group is in the first mover position. The adverse effect, however, is avoided if the smaller group demonstrates unwillingness to form a coalition or if it has high costs in producing the public good. Hence, another paradoxical effect in global public good provision arises: Green technological progress, which enables the small coalition to make an effective contribution to the public good, may in the end reduce public good supply and thus aggravate the underprovision problem.

This additional feature was not considered in this paper [see Buchholz and Eichenseer (2016), for an elaboration on this]. For the sake of simplicity, it had rather been assumed that production costs for the public good are exogenously given and identical for all countries. Moreover, there have only been two potential coalitions whose members have the same endowments and preferences. In further research, one might drop this assumption and allow for a greater number of heterogeneous groups whose members can differ w.r.t. their public good productivities their income levels and their preferences.

A.1 Appendix: Derivation of the Threshold Level in Proposition 11

It directly follows from Eq. (19) that

$$\frac{\partial G_I^B}{\partial \xi} = \frac{w - e_K - \xi e_2^K}{1 + \xi e_1^K + m e_1^M} \tag{29}$$

which is positive since $e_1^K > 0$, $e_1^M > 0$, $e_2^K < 0$ and $w - e^K$ is a member of K's private consumption which is positive in the equilibrium $E_I^B(k, m)$ by definition. Inserting (29) into

$$\frac{\partial u_I^{BK}}{\partial \xi} = \frac{\partial u\left(e\left(G_I^B, \xi\right), G_I^B\right)}{\partial \xi} = u_2\left(\left(\xi e_1^K + 1\right)\frac{\partial G_I^B}{\partial \xi} + \xi e_2^K\right) \tag{30}$$

gives

$$\frac{\partial u_I^{BK}}{\partial \xi} = \frac{u_2\left(\left(\xi e_1^K + 1\right)(w - e_K) + \xi m e_2^K e_1^M\right)}{1 + \xi e_1^K + m e_1^M}. \tag{31}$$

Then an upper bound $\underline{\chi} > \underline{\xi}_B(m)$ which has the properties required by Proposition 11 exist since at $\xi = \underline{\xi}_B(m)$ we have $w - e^K = 0$. Hence, because of $e_1^M > 0$ and $e_2^K < 0$, the numerator is of (30) is negative there so that continuity implies that u_I^{BK} must also be falling for all ξ close to $\underline{k}_B(m)$. QED

References

Andreoni, J. (1988). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics, 35*(1), 57–73.

Andreoni, J., & McGuire, M. C. (1993). Identifying the free riders: A simple algorithm for determining who will contribute to a public good. *Journal of Public Economics, 51*(3), 447–454.

Bergstrom, T. C., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics, 29*(1), 25–49.

Boadway, R., & Hayashi, M. (1999). Country size and the voluntary provision of international public goods. *European Journal of Political Economy, 15*(4), 619–638.

Buchholz, W., Cornes, R., & Rübbelke, D. (2014). Potentially harmful international cooperation on global public good provision. *Economica, 81*(2), 205–223.

Buchholz, W., & Eichenseer, M. (2016). *How green technological progress can be bad for the environment*. Discussion Paper, University of Regensburg

Buchholz, W., Haslbeck, C., & Sandler, T. (1998). When does partial cooperation pay? *Finanzarchiv, 55*, 1–20.

Buchholz, W., Haupt, A., & Peters, W. (2013). International environmental agreements, fiscal federalism, and constitutional design. *Review of International Economics, 21*(4), 705–718.

Buchholz, W., & Sandler, T. (2016). Successful leadership in global public good provision: Incorporating behavioural approaches. *Environmental and Resource Economics*, 1–17. doi:10.1007/s10640-016-9997-2.

Chen, C., & Zeckhauser, R. (2016). *Collective action in an asymmetric world* (NBER Working Paper No. 22240).

Cornes, R. (2016). Aggregative environmental games. *Environmental and Resource Economics, 63*(2), 339–365.

Cornes, R., & Hartley, R. (2007). Aggregative public good games. *Journal of Public Economic Theory, 9*(2), 201–219.

Cornes, R., & Sandler, T. (1986/1996). *The theory of externalities, public goods, and club goods* (1st ed./2nd ed.). Cambridge: Cambridge University Press.

Eckert, H. (2003). Negotiating environmental agreements: Regional or federal authority? *Journal of Environmental Economics and Management, 46*(1), 1–24.

Foucart, R., & Wan, C. (2016). *Strategic decentralization and the provision of global public goods.* Discussion Paper, Humboldt University Berlin.

Hattori, K. (2015). Within-group cooperation and between-group externalities in the provision of public goods. *International Tax and Public Finance, 22*(2), 252–273.

Hoel, M. (1991). Global environmental problems: The effects of unilateral action taken by one country. *Journal of Environmental Economics and Management, 20*(1), 55–70.

Lindahl, E. (1919). *Die Gerechtigkeit der Besteuerung. Eine Analyse der Steuerprincipien auf Grundlage der Grenznutzentheorie*. Lund: Gleerupska Universitetsbokhandeln.

Peinhardt, C., & Sandler, T. (2015). *Transnational cooperation: An issue-based approach*. Oxford UK: Oxford University Press.

Sandler, T. (2004). *Global collective action*. Cambridge, MA: Cambridge University Press.

Sandler, T. (2016). *Environmental cooperation: Contrasting international environmental agreements.* Discussion Paper, University of Texas at Dallas.

Sinn, H.-W. (2012). *The green paradox*. Cambridge, MA: The MIT Press.

Vicary, S. (2016). *Is three a crowd? Small group provision of a public good.* Discussion Paper, University of Newcastle-upon-Tyne.

Wicksell, K. (1896). *Finanztheoretische Untersuchungen*. Jena: Gustav Fischer.

Evolution of Consistent Conjectures in Semi-aggregative Representation of Games, with Applications to Public Good Games and Contests

Alex Possajennikov

1 Introduction

Aggregative games are a special class of games in which the payoff of a player depends on the player's own strategy and on a (common across players) aggregate of players' strategies. An early example of a work on aggregative games is Corchón (1994) but in a more recent series of works, Richard Cornes and Roger Hartley elucidated the usefulness of studying the mathematical structure of these games for establishing equilibrium existence and for finding equilibria in situations going beyond textbook symmetric examples. They applied this methodology to such classic examples of economic analysis as public good games (Cornes and Hartley 2007) and contests (Cornes and Hartley 2003, 2005),[1] as well as studying the general structure of aggregate games further (Cornes and Hartley 2012).

Before turning his attention to aggregative games, Richard also worked on applications of the concept of conjectural variations. This concept was extensively analyzed in the context of industrial organization games (see e.g. Laitner 1980; Bresnahan 1981; Perry 1982); its application in common property exploitation model was considered in Cornes and Sandler (1983) and in public good games in Cornes and Sandler (1984a).

Paper prepared for a volume honoring the memory of Richard Cornes. In his time at the University of Nottingham, Richard was a helpful colleague, ready to give advice in his usual witty and entertaining manner.

[1] Further examples of aggregative games are listed in Cornes and Hartley (2011) and Cornes (2016) discusses the applications of aggregative games in the analysis of environmental problems.

A. Possajennikov (✉)
School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK
e-mail: alex.possajennikov@nottingham.ac.uk

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_5

In this paper, I also focus on a representation of games that is similar to the aggregative one and on conjectural variations. The representation is such that a player's payoff depends on the player's strategy and on a certain aggregate of all player's strategies, personalized for the player. Thus the aggregates do not have to be the same for all players, as in a usual aggregative game. Nevertheless, the aggregates fulfill a similar role of reducing the dimensionality of what a player needs to consider about the other players. In such a representation (which I call *semi-aggregative*), what is relevant for the player is how the aggregate measure of players' strategies possibly changes. This is precisely what the conjectures of the players in such a game are about.

Given their conjectures about possible changes in the respective aggregates, the players in the game behave rationally, that is, maximize their payoff. Their decisions characterize an equilibrium, for the given conjectures. But where do the conjectures come from? I suppose that they represent players' innate beliefs, but those beliefs are subject to evolution. Different conjectures will lead to different equilibrium choices and thus different payoffs. From the point of view of evolution, those conjectures that led to higher payoff are more likely to propagate.[2]

I focus on the setting where a game is not necessarily symmetric, thus players can have different roles (for example, one player can have a larger marginal benefit from a public good than another player, or a lower cost of contributing to it). Since roles are different, evolution is considered as happening within each role separately. Instead of considering an explicit dynamic process, I look for evolutionarily stable conjectures, which are conjectures that no other conjectures can invade by achieving a higher payoff for this player's role, given the conjectures of the other players and the equilibrium that the players play.

I find that the evolutionary stability of conjectures is linked to their consistency. An equilibrium in the model is at the intersection of the reaction functions of the players, which also define the reaction of the aggregates. If a player's strategy changes, for whatever reason, the reaction functions determine how the other players change (optimally) their strategies, and thus how the aggregates change. Conjectures are considered consistent if the belief of a player locally coincides, to the first approximation, to the actual change in the player's personalized aggregate. The main result of the paper is that, in well-behaved games, only consistent conjectures of a player can be evolutionary stable for this player.

The result extends the link between consistent and evolutionarily stable conjectures. Previous works noted this connection in simple duopoly models (Dixon and Somma 2003; Müller and Normann 2005), in two-player games (Possajennikov 2009) and in symmetric aggregative games (Possajennikov 2015). What I add in this paper is that the link between consistent and evolutionarily stable conjectures hold

[2]Another interpretation is that players first choose conjectures and then play the game. The search is then for an equilibrium in the game of choosing conjectures. I nevertheless prefer the evolutionary interpretation, which makes it clearer that the process of forming beliefs and choosing strategies occur at different times. This evolutionary interpretation is an example of the "indirect evolution approach" (Güth and Yaari 1992).

in more general n-player asymmetric situations. Thus it is not only that evolution selects consistent conjectures when other players' conjectures are consistent; for any conjectures of the other players, it is best, from the evolutionary point of view, to have a consistent conjecture.

This result is illustrated on two examples of games that were often the subject of Richard Cornes's work and that are aggregative or naturally semi-aggregative, namely public good games and contests. In these settings, I show that for many parameter values consistent and evolutionarily stable conjectures coincide, thus consistency is not only a necessary but also a sufficient condition for evolutionary stability.

2 Games and Conjectures

2.1 Semi-aggregative Representation of Games

A simultaneous-move game on the real line is $G = (N, \{X_i\}_{i=1}^n, \{u_i\}_{i=1}^n)$, where $N = \{1, \ldots, n\}$ is the set of players, $X_i \subset \mathbb{R}$ is the strategy set of player i, and $u_i : X_1 \times \ldots \times X_n \to \mathbb{R}$ is the payoff function of player i. It is assumed that the game is well-behaved: strategy sets are convex and the payoff functions are differentiable as many times as required.

For any game, the payoff of player i can be written as $u_i(x_i, A_i)$, where $A_i = f_i(x_1, \ldots, x_n)$ for some function $f_i : X_1 \times \ldots \times X_n \to \mathbb{R}$.[3] I call the representation of the payoffs in the form $u_i(x_i, A_i)$ *semi-aggregative*, since A_i can be seen as a personalized aggregate of player i, which summarizes the dependence of the payoff of player i on the strategies of other players. Note that the aggregate A_i can include the strategy x_i of player i. A game is *aggregative* if there exists a semi-aggregative representation with $A = A_i$ for all i, i.e. with the same functions f_i for all players and a common aggregate A.

While in general games the payoff representation discussed above may appear strange, there are classes of games for which a (semi-)aggregative representation is natural. For example, in a differentiated product oligopoly, the price $p_i(q_1, \ldots, q_n)$ for the product of firm i is determined by the inverse demand from the quantities chosen by all firms. This price can then naturally be taken as the personalized aggregate of firm i. The payoff for firm i is the profit $\pi_i(q_i, p_i) = p_i q_i - C_i(q_i)$, where $C_i(q_i)$ is the cost function of firm i.[4]

For another example, consider a (pure) public good game. Each player i contributes a part x_i of the endowment m_i to the public good, leaving $m_i - x_i$

[3]For example, consider the identity $u_i(x_1, \ldots, x_n) = x_i + u_i(x_1, \ldots, x_n) - x_i$. Let $A_i = f_i(x_1, \ldots, x_n) = u_i(x_1, \ldots, x_n) - x_i$ and write $u_i(x_i, A_i) = x_i + A_i$.

[4]In a homogeneous good market with one price, the aggregate (the price) is the same for all firms, and the game is properly aggregative.

for private good consumption. The aggregate production of the public good is $A = \sum_{i=1}^{n} x_i$. Player i's payoff is given by the utility function $u_i(m_i - x_i, A)$, which is already a semi-aggregative representation. In fact, with pure public goods the game is properly aggregative, since the aggregate amount of public good A is the same for all players; it is not needed to have personalized aggregates for each player.

The advantage of the semi-aggregative representation is the reduction in the dimensionality of the problem. In a sense, a player sees his or her opponents as one aggregate opponent and is only concerned about the aggregate effect of such an opponent on payoff. In the next section I discuss how this can be used to formulate in a simple manner players' expectations about the behavior of other players.

2.2 Conjectures and Conjectural Variation Equilibria

Suppose that player i has some conjectures r_i about the reaction of other players to a change in the player's own strategy. With the semi-aggregative representation of the game, the conjectures are about the change in the personalized aggregate, $r_i = \left(\frac{dA_i}{dx_i}\right)^e$, where the superindex is meant to convey that it is an *expected*, rather than actual, change. It is assumed that the conjectures are constant, $r_i \in R_i$, where R_i is a convex subset of $\mathbb{R}$, i.e. conjectures do not depend on the current strategies of players. This assumption again reduces the dimensionality of the problem while still allowing consideration of consistent conjectures.

A change in player i's own strategy x_i also can directly affect the aggregate $A_i = f_i(x_1, \ldots, x_n)$. But the conjecture is about the *total* effect of a change in x_i on A_i: it incorporates the direct effect $\frac{\partial A_i}{\partial x_i}$ but also the effect from the expected changes in the other players' strategies. This formulation is slightly more general than the one with the aggregate being a function of the other players' strategies only, as was used in e.g. Perry (1982) for oligopoly and in Cornes and Sandler (1984a) for a public good model. It can still represent the usual Nash behavior: $r_i = \frac{\partial A_i}{\partial x_i}$ means that the strategies of the other players are kept fixed; player i does not expect the other players to react.

Having conjecture r_i, player i maximizes payoff $u_i(x_i, A_i)$. The first-order condition for maximization is

$$F_i(x_i, A_i; r_i) = \frac{\partial u_i}{\partial x_i}(x_i, A_i) + \frac{\partial u_i}{\partial A_i}(x_i, A_i) \cdot r_i = 0. \tag{1}$$

Suppose now that all players have certain conjectures, summarized by vector $\mathbf{r} = (r_1, \ldots, r_n)$. Suppose further that for each player i, the solution of the player's maximization problem is characterized by Eq. (1). A *conjectural variation equilibrium (CVE)* for the given vector $\mathbf{r}$ of conjectures consists of the vector of players' strategies $\mathbf{x}^*(\mathbf{r}) = (x_1^*(\mathbf{r}), \ldots, x_n^*(\mathbf{r}))$ and the vector of personalized aggregates $\mathbf{A}^*(\mathbf{r}) = (A_1^*(\mathbf{r}), \ldots, A_n^*(\mathbf{r}))$ that satisfy the system of equations

$$F_i(x_i, A_i; r_i) = 0,\ i = 1, \ldots, n, \qquad (2)$$
$$A_i - f_i(x_i, \ldots, x_n) = 0,\ i = 1, \ldots, n.$$

It is assumed that the solution of this system of equations exists for the values of conjectures in sets R_i. There may be multiple solutions of the system; in the analysis below I consider any particular solution that is locally unique and well-behaved.

Although conjectures are about changes in a player's strategy and reactions to them, the conjectural variation equilibrium is a static concept. However, it can be interpreted as a convenient short-cut summarizing the result of a more explicit dynamic analysis,[5] and this is the interpretation I have in mind by focusing on CVE in this paper.

2.3 Consistent Conjectures

Recall that a conjecture of player i is a belief about the change in the personalized aggregate A_i in response to a change in player i's strategy x_i. To define consistent conjectures, let x_i vary unconstrained and concentrate on optimal responses of the other players. Consider the system of equations

$$F_j(x_j, A_j; r_j) = 0, j = 1, \ldots, n, j \neq i \qquad (3)$$
$$A_j - f_j(x_1, \ldots, x_n) = 0, \quad j = 1, \ldots, n,$$

which is like system (2) except that the first-order condition for player i is not there. Thus, the strategy x_i of player i is not constrained to be optimal; it can take any value. The strategies of the other players are still characterized by the first-order conditions; thus the system describes optimal responses of the other players to arbitrary values of x_i. Denote a solution of system (3) as $(x_1^{**}(x_i), \ldots, x_{i-1}^{**}(x_i), x_{i+1}^{**}(x_i), \ldots, x_n^{**}(x_i); A_1^{**}(x_i), \ldots, A_n^{**}(x_i))$.

Consider a vector of conjectures $\mathbf{r}$ and a certain CVE $(\mathbf{x}^*, \mathbf{A}^*) = (x_1^*, \ldots, x_n^*; A_1^*, \ldots, A_n^*)$ for these conjectures. Note that for $x_i = x_i^*$ there exists a solution of system (3) with $x_j^{**}(x_i^*) = x_j^*$ for all $j \neq i$ and $A_j^{**}(x_i^*) = A_j^*$ for all $j = 1, \ldots, n$. Consider such a solution and consider $A_i^{**}(x_i)$. Conjecture r_i of player

[5] In a duopoly context, Dockner (1992) and Cabral (1995) show that a dynamic model indeed can lead to the same outcomes as certain CVEs, and Itaya and Dasgupta (1995) and Itaya and Okamura (2003) do so for a public good game.

i is *consistent* if $r_i = \frac{dA_i^{**}}{dx_i}(\mathbf{x}^*, \mathbf{A}^*)$, i.e. the conjecture about the reaction of the personalized aggregate is, to a first approximation, correct at equilibrium.[6]

Whether a particular conjecture r_i is consistent depends on the vector of conjectures $\mathbf{r}_{-i}$ of the other players. Given a vector $\mathbf{r}$, it is possible that some players hold consistent conjectures and others not. One can define conjectures to be mutually consistent if for all i, r_i is consistent against $\mathbf{r}_{-i}$. However, it will not be important for the analysis of conjectures of player i what conjectures the other players hold thus I do not focus only on mutually consistent conjectures.

3 Evolutionary Stability of Conjectures

Imagine that for each of the n player roles there is a large (infinite) population of players, and players from each population from time to time are called to play the game G against opponents randomly drawn from the other populations. Consider the population for the role of player i. Each player in the population has some conjectures. Suppose that in all other player populations conjectures have stabilized on some values $\mathbf{r}_{-i}$. Thus, if called to play, a player with a certain conjecture r_i from the population of players i will play the game against the other players with conjectures $\mathbf{r}_{-i}$. Suppose that when the game is played, a CVE is played. The question is: for the given conjectures $\mathbf{r}_{-i}$ of the other players, which conjecture of player i is evolutionarily stable?

Different conjectures in the population for the role of player i will lead to different CVEs and thus to different payoffs. Conjecture r_i^{ES} is said to be *evolutionarily stable* (Maynard Smith and Price 1973; Selten 1980) if

$$u_i(x_i^*(r_i^{ES}, \mathbf{r}_{-i}), A_i^*(r_i^{ES}, \mathbf{r}_{-i})) > u_i(x_i^*(r_i, \mathbf{r}_{-i}), A_i^*(r_i, \mathbf{r}_{-i})) \text{ for any } r_i \neq r_i^{ES}.$$

The above inequality means that in the population for the role of player i, a player with conjecture r_i^{ES} will get a higher payoff when called to play than a player with any other value r_i of the conjecture. The evolutionary intuition is that players with any other conjecture r_i in the population for the role of player i would have lower fitness than the players with conjecture r_i^{ES}. Therefore evolution will favor players with conjecture r_i^{ES} to survive and thrive.[7]

With the alternative interpretation that players first choose their conjectures and then play a CVE of the game G, an evolutionarily stable conjecture of player i is a strict best response of player i to the given conjectures of the other players. If a vector of conjectures $\mathbf{r}^{ES} = (r_1^{ES}, \dots, r_n^{ES})$ is such that for each player i the

[6]This consistency requirement was introduced by Bresnahan (1981) for a duopoly, and also used e.g. in Perry (1982) in an oligopoly and Cornes and Sandler (1984a) in a public good game context.

[7]Note that the definition focuses on player i treating the other players conjectures as fixed; Selten (1980) showed that such an approach is appropriate in asymmetric games.

conjecture r_i^{ES} is evolutionarily stable given $\mathbf{r}_{-i}^{ES}$, then $(r_1^{ES}, \ldots, r_n^{ES})$ is a strict Nash equilibrium in the game where players choose conjectures and their payoffs are determined via conjectural variations equilibria.

Whatever the interpretation, an evolutionary stable conjecture solves

$$\max_{r_i} u_i(x_i^*(r_i, \mathbf{r}_{-i}), A_i^*(r_i, \mathbf{r}_{-i})).$$

The first-order condition for maximization is[8]

$$\frac{\partial u_i}{\partial x_i}\frac{\partial x_i^*}{\partial r_i} + \frac{\partial u_i}{\partial A_i}\frac{dA_i^*}{dr_i} = 0.$$

Therefore (provided that $\frac{\partial u_i}{\partial A_i} \neq 0$ and $\frac{\partial x_i^*}{\partial r_i} \neq 0$), $-\frac{\partial u_i/\partial x_i}{\partial u_i/\partial A_i} = \frac{dA_i^*/dr_i}{\partial x_i^*/\partial r_i}$. Since from Eq. (1) $r_i = -\frac{\partial u_i/\partial x_i}{\partial u_i/\partial A_i}$, an interior evolutionarily stable conjecture satisfies

$$r_i^{ES} = \frac{dA_i^*/dr_i}{\partial x_i^*/\partial r_i}. \tag{4}$$

Speaking somewhat loosely in mathematical terms, if $dr_i = \partial r_i$ is treated as a small change in the independent variable r_i, then it can be canceled from (4). Note also that $dx_i = dx_i^* = \partial x_i^*$ if only r_i changes. Therefore $r_i^{ES} = \frac{dA_i^*}{dx_i}$. Recall that a conjecture is consistent if $r_i = \frac{dA_i^{**}}{dx_i}$. Since at a CVE $A_i^{**}(x_i^*) = A_i^*$, the first-order condition for evolutionary stability and the consistency condition are essentially the same.[9]

For a more formal demonstration of the reasoning, consider system (2). To simplify notation, focus on $i = 1$. Differentiating each line of (2) with respect to r_1,

$$\begin{array}{ccccccccc}
\frac{\partial F_1}{\partial x_1}\frac{\partial x_1^*}{\partial r_1} & + \ldots + & 0 & + & \frac{\partial F_1}{\partial A_1}\frac{dA_1^*}{dr_1} & + \ldots + & 0 & = & -\frac{\partial F_1}{\partial r_1} \\
\cdots & \cdots & \cdots & & \cdots & \cdots & & & \cdots \\
0 & + \ldots + & \frac{\partial F_n}{\partial x_n}\frac{\partial x_n^*}{\partial r_1} & + & 0 & + \ldots + & \frac{\partial F_n}{\partial A_n}\frac{dA_n^*}{dr_1} & = & 0 \\
-\frac{\partial f_1}{\partial x_1}\frac{\partial x_1^*}{\partial r_1} & + \ldots + & -\frac{\partial f_1}{\partial x_n}\frac{\partial x_n^*}{\partial r_1} & + & \frac{dA_1^*}{dr_1} & + \ldots + & 0 & = & 0 \\
\cdots & \cdots & \cdots & & \cdots & \cdots & & & \cdots \\
-\frac{\partial f_n}{\partial x_1}\frac{\partial x_1^*}{\partial r_1} & + \ldots + & -\frac{\partial f_n}{\partial x_n}\frac{\partial x_n^*}{\partial r_1} & + & 0 & + \ldots + & \frac{dA_n^*}{dr_1} & = & 0
\end{array}$$

[8]To save space, arguments of derivatives are omitted. It is understood that they are evaluated at $\mathbf{r} = (r_i^{ES}, \mathbf{r}_{-i})$ and CVE $(\mathbf{x}^*(\mathbf{r}), \mathbf{A}^*(\mathbf{r}))$.

[9]The relationship between consistent conjectures and the conjectures that maximize the indirect payoff function $u_i(x_i^*(r_i, r_j), A_i^*(r_i, r_j))$ was noted by Itaya and Dasgupta (1995) for a two-player public good game.

Define

$$M = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \dots & 0 & \frac{\partial F_1}{\partial X_1} & \dots & 0 \\ \dots & \ddots & \dots & \dots & \ddots & \dots \\ 0 & \dots & \frac{\partial F_n}{\partial x_n} & 0 & \dots & \frac{\partial F_n}{\partial X_n} \\ -\frac{\partial f_1}{\partial x_1} & \dots & -\frac{\partial f_1}{\partial x_n} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \dots & \ddots & \dots \\ -\frac{\partial f_n}{\partial x_1} & \dots & -\frac{\partial f_n}{\partial x_n} & 0 & \dots & 1 \end{pmatrix},$$

$$M_{-11} = \begin{pmatrix} \frac{\partial F_2}{\partial x_2} & \dots & 0 & \frac{\partial F_2}{\partial X_2} & \dots & 0 \\ \dots & \ddots & \dots & \dots & \ddots & \dots \\ 0 & \dots & \frac{\partial F_n}{\partial x_n} & 0 & \dots & \frac{\partial F_n}{\partial X_n} \\ -\frac{\partial f_2}{\partial x_2} & \dots & -\frac{\partial f_2}{\partial x_n} & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \dots & \ddots & \dots \\ -\frac{\partial f_n}{\partial x_2} & \dots & -\frac{\partial f_n}{\partial x_n} & 0 & \dots & 1 \end{pmatrix},$$

and

$$M_{-1A} = \begin{pmatrix} 0 & \frac{\partial F_2}{\partial x_2} & \dots & 0 & \frac{\partial F_2}{\partial X_2} & \dots & 0 \\ \vdots & \dots & \ddots & \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \frac{\partial F_n}{\partial x_n} & 0 & \dots & \frac{\partial F_n}{\partial X_n} \\ -\frac{\partial f_1}{\partial x_1} & -\frac{\partial f_1}{\partial x_2} & \dots & -\frac{\partial f_1}{\partial x_n} & 0 & \dots & 0 \\ -\frac{\partial f_2}{\partial x_1} & -\frac{\partial f_2}{\partial x_2} & \dots & -\frac{\partial f_2}{\partial x_n} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \dots & \ddots & \dots \\ -\frac{\partial f_n}{\partial x_1} & -\frac{\partial f_n}{\partial x_2} & \dots & -\frac{\partial f_n}{\partial x_n} & 0 & \dots & 1 \end{pmatrix}.$$

If $|M| \neq 0$, by Cramer's rule, $\frac{\partial x_1^*}{\partial r_1} = \frac{1}{|M|}\left(-\frac{\partial F_1}{\partial r_1}\right)|M_{-11}|$ and $\frac{\partial A_1^*}{\partial r_1} = \frac{1}{|M|}(-1)^n \frac{\partial F_1}{\partial r_1}|M_{-1A}|$. Therefore, if $|M_{-11}| \neq 0$ (from Eq. (1) $\frac{\partial F_1}{\partial r_1} = \frac{\partial u_i}{\partial A_i}$ thus $\frac{\partial F_1}{\partial r_1} \neq 0$ if $\frac{\partial u_i}{\partial A_i} \neq 0$), Eq. (4) becomes

$$r_1^{ES} = \frac{(-1)^{n-1}|M_{-1A}|}{|M_{-11}|}. \tag{5}$$

To determine $\frac{dA_1^{**}}{dx_1}$, consider system (3). Differentiating each of the equations with respect to x_1,

$$\begin{array}{cccccccccc}
\frac{\partial F_2}{\partial x_2}\frac{\partial x_2^{**}}{\partial x_1} & +\ldots+ & 0 & + & 0 & + & \frac{\partial F_2}{\partial A_2}\frac{dA_2^{**}}{dx_1} & +\ldots+ & 0 & = & 0 \\
\ldots & \ldots & \ldots & & \ldots & & \ldots & \ldots & \ldots & & \ldots \\
0 & +\ldots+ & \frac{\partial F_n}{\partial x_n}\frac{\partial x_n^{**}}{\partial x_1} & + & 0 & + & 0 & +\ldots+ & \frac{\partial F_n}{\partial A_n}\frac{dA_n^{**}}{dx_1} & = & 0 \\
-\frac{\partial f_1}{\partial x_2}\frac{\partial x_2^{**}}{\partial x_1} & +\ldots+ & -\frac{\partial f_1}{\partial x_n}\frac{\partial x_n^{**}}{\partial x_1} & + & \frac{dA_1^{**}}{dx_1} & + & 0 & +\ldots+ & 0 & = & \frac{\partial f_1}{\partial x_1} \\
\ldots & \ldots & \ldots & & \ldots & & \ldots & \ldots & \ldots & & \ldots \\
-\frac{\partial f_n}{\partial x_2}\frac{\partial x_2^{**}}{\partial x_1} & +\ldots+ & -\frac{\partial f_n}{\partial x_n}\frac{\partial x_n^{**}}{\partial x_1} & + & 0 & + & 0 & +\ldots+ & \frac{dA_n^{**}}{dx_1} & = & \frac{\partial f_n}{\partial x_1}
\end{array}$$

Define

$$L_{-11} = \begin{pmatrix}
0 & \frac{\partial F_2}{\partial x_2} & \ldots & 0 & \frac{\partial F_2}{\partial A_2} & \ldots & 0 \\
\vdots & \ldots & \ddots & \ldots & \ldots & \ddots & \ldots \\
0 & 0 & \ldots & \frac{\partial F_n}{\partial x_n} & 0 & \ldots & \frac{\partial F_n}{\partial A_n} \\
1 & -\frac{\partial f_1}{\partial x_2} & \ldots & -\frac{\partial f_1}{\partial x_n} & 0 & \ldots & 0 \\
0 & -\frac{\partial f_2}{\partial x_2} & \ldots & -\frac{\partial f_2}{\partial x_n} & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \ldots & \ddots & \ldots \\
0 & -\frac{\partial f_n}{\partial x_2} & \ldots & -\frac{\partial f_n}{\partial x_n} & 0 & \ldots & 1
\end{pmatrix}$$

and

$$L_{-1A} = \begin{pmatrix}
0 & \frac{\partial F_2}{\partial x_2} & \ldots & 0 & \frac{\partial F_2}{\partial A_2} & \ldots & 0 \\
\vdots & \ldots & \ddots & \ldots & \ldots & \ddots & \ldots \\
0 & 0 & \ldots & \frac{\partial F_n}{\partial x_n} & 0 & \ldots & \frac{\partial F_n}{\partial A_n} \\
\frac{\partial f_1}{\partial x_1} & -\frac{\partial f_1}{\partial x_2} & \ldots & -\frac{\partial f_1}{\partial x_n} & 0 & \ldots & 0 \\
\frac{\partial f_2}{\partial x_1} & -\frac{\partial f_2}{\partial x_2} & \ldots & -\frac{\partial f_2}{\partial x_n} & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ldots \\
\frac{\partial f_n}{\partial x_1} & -\frac{\partial f_n}{\partial x_2} & \ldots & -\frac{\partial f_n}{\partial x_n} & 0 & \ldots & 1
\end{pmatrix}$$

Then $|L_{-11}| = (-1)^n|M_{-11}|$ and $|L_{-1A}| = -|M_{-1A}|$. Using Cramer's rule again, $\frac{dA_1^{**}}{dx_1} = \frac{1}{|L_{-11}|}|L_{-1A}| = \frac{-|M_{-1A}|}{(-1)^n|M_{-11}|} = \frac{(-1)^{n-1}|M_{-1A}|}{|M_{-11}|}$, which is the same as the right-hand side of Eq. (5). Thus, the following proposition is proved:

Proposition 1 *Consider a semi-aggregative representation of the game G and consider conjecture profile* $\mathbf{r} = (r_1, \ldots, r_n)$. *Suppose that there exists a CVE* $(\mathbf{x}^*(\mathbf{r}), \mathbf{A}^*(\mathbf{r}))$ *for this* $\mathbf{r}$. *If* $\frac{\partial u_i}{\partial A_i} \neq 0$, $|M| \neq 0$ *and* $|M_{-11}| \neq 0$ *at* $\mathbf{r}$ *and*

$(\mathbf{x}^*(\mathbf{r}), \mathbf{A}^*(\mathbf{r}))$*, then if* r_i *is an evolutionarily stable conjecture for player i, then it is a consistent conjecture for player i.*

Since the analysis was based only on the first-order condition for evolutionary stability, it is not necessarily the case that a consistent conjecture is evolutionarily stable. Concavity or quasi-concavity conditions on the indirect function $u_i(x_i(\mathbf{r}), A_i(\mathbf{r}))$ can guarantee this. Instead of stating these conditions in general, evolutionarily stability of consistent conjectures is demonstrated for particular games in Sect. 4.

The practical usefulness of the result is that it is usually easier to find consistent conjectures than to derive the indirect function to search for the evolutionarily stable ones. Since the result shows that in well-behaved games only consistent conjectures can be evolutionarily stable in the interior of the conjecture space, the search for evolutionarily stable conjectures can be reduced to the consistent ones.

The conceptual usefulness of the result is to provide foundations for consistent conjectures. Consistency of conjectures is not always accepted as a plausible criterion for preferring some conjectures over others.[10] The result in this paper shows though, that if players are endowed with conjectures that are subject to evolutionary pressure (or, equivalently, if players could choose conjectures before playing the game), then only consistent conjectures can survive such a process.

Note that the proof of the result concentrated on player i, while taking arbitrary conjectures held by the other players. The conjectures of the other players may or may not be consistent; if one wants all players to have evolutionarily stable conjectures, then only profiles with mutually consistent conjectures can be such. The result shows that it is best for player i to have consistent conjectures whatever the conjectures of the other players are (but which value of the conjecture is consistent, and thus possibly evolutionarily stable, for player i depends on the current conjectures of the other players).[11]

The current result generalizes the previous ones in Possajennikov (2009, 2015) to asymmetric games with more than two players. In principle, the games do not even need to have an obvious aggregative structure: what was used is that the players make conjectures about the appropriate quantity A_i that was relevant for their payoff. In general, the function f_i determining this quantity may be complicated and thus it is not likely that the players would consider conjectures about it; however, in some games, illustrated in the next section, the aggregate quantity A_i arises naturally in the formulation of the problem.

[10] See e.g. Makowski (1987) and Cornes and Sandler (1996, p. 32) say that they do not attach any particular importance to consistent conjectures.

[11] The observation that the consistent conjecture is the best response conjecture of one player to any given conjecture of the other player was made in Dixon and Somma (2003) for a linear-quadratic Cournot duopoly game.

4 Examples

4.1 Semi-public Good Games

Cornes and Sandler (1984a,b) explored the public good model, including the impact of various conjectures and the possibility of impure public goods, where a player's contribution to a public good also provides a private benefit. I will use instead the formulation of semi-public goods from Costrell (1991) that models the same idea—that a player benefits more from his or her own contribution to a public good than the other players do—in a more transparent manner. The formulation also encompasses a pure public good model.

Suppose that each player i has a money endowment m_i that can be spent either on a private good or on a semi-public good. Assuming for simplicity that prices of all goods are equal and normalizing the price to 1, $m_i = y_i + x_i$, where y_i is the amount spent on the private good and x_i the amount spent on the public good. Player i has the utility function $u_i(y_i, G_i)$, where G_i is the quantity of the public good available to player i. The semi-public nature of the public good is modeled by $G_i = x_i + b_i \sum_{j \neq i} x_j$, where $0 < b_i \leq 1$. Player i benefits most from his or her own contribution to the public good, but other players' contributions also spillover to player i's benefit. Quantity G_i naturally plays the role of the personalized aggregate for player i.[12]

To illustrate the result in the previous section, consider the three-player case ($n = 3$) and Cobb-Douglas utility functions for all players

$$u_i(x_i, G_i) = (m_i - x_i)^{\alpha_i} G_i^{1-\alpha_i},$$

with $0 < \alpha_i < 1$. Suppose that each player i has conjecture $r_i \geq 0$. Player i's first-order condition for utility maximization is $-\alpha_i(m_i - x_i)^{\alpha_i - 1} G_i^{1-\alpha_i} + (1 - \alpha_i)(m_i - x_i)^{\alpha_i} G_i^{-\alpha_i} r_i = 0$, or, in the interior where $x_i \neq m_i$ and $G_i \neq 0$, $-\alpha_i G_i + (1 - \alpha_i)(m_i - x_i) r_i = 0$. Therefore

$$\alpha_i G_i + (1 - \alpha_i) x_i r_i = (1 - \alpha_i) m_i r_i$$

characterizes the solution of player i utility maximization problem.[13]

To find consistent conjectures for player 1, consider the system

$$\alpha_2 G_2 + (1 - \alpha_2) x_2 r_2 = (1 - \alpha_2) m_2 r_2$$
$$\alpha_3 G_3 + (1 - \alpha_3) x_3 r_3 = (1 - \alpha_3) m_3 r_3$$

[12] Note that if $b_i = 1$ for all i, then the public good becomes a pure public good and the same aggregate $G = \sum_{i=1}^{n} x_i$ can be used for all players.

[13] The second-order condition $\alpha_i(1-\alpha_i)(m_i-x_i)^{\alpha_i-2} G_i^{-\alpha_i-1}(-G_i^2 - 2(m_i-x_i)G_i r_i - (m_i-x_i)^2 r_i^2) < 0$ is satisfied for $r_i \geq 0$ and all interior x_i, G_i.

$$G_1 - x_1 - b_1 x_2 - b_1 x_3 = 0$$
$$G_2 - b_2 x_1 - x_2 - b_2 x_3 = 0$$
$$G_3 - b_3 x_1 - b_3 x_2 - x_3 = 0.$$

Substituting the last three equations into the first two,

$$((1-\alpha_2)r_2 + \alpha_2)x_2 + \alpha_3 b_3 x_3 = (1-\alpha_2)m_2 r_2 - \alpha_2 b_2 x_1$$
$$\alpha_2 b_2 x_2 + ((1-\alpha_3)r_3 + \alpha_3)x_3 = (1-\alpha_3)m_3 r_3 - \alpha_3 b_3 x_1.$$

If $b_j < 1, j = 2, 3$, then the solution of these two equations is guaranteed to exist. It is

$$x_2^{**}(x_1) = \frac{(m_2(1-\alpha_2)r_2 - \alpha_2 b_2 x_1)((1-\alpha_3)r_3 + \alpha_3) - \alpha_2 b_2 (m_2(1-\alpha_3)r_3 - \alpha_3 b_3 x_1)}{((1-\alpha_2)r_2 + \alpha_2)((1-\alpha_3)r_3 + \alpha_3) - \alpha_2 b_2 \alpha_3 b_3}$$
$$x_3^{**}(x_1) = \frac{((1-\alpha_2)r_2 + \alpha_2)(m_3(1-\alpha_3)r_3 - \alpha_3 b_3 x_1) - (m_2(1-\alpha_2)r_2 - \alpha_2 b_2 x_1)\alpha_3 b_3}{((1-\alpha_2)r_2 + \alpha_2)((1-\alpha_3)r_3 + \alpha_3) - \alpha_2 b_2 \alpha_3 b_3}.$$

Since $G_1^{**}(x_1) = x_1 + b_1 x_2^{**}(x_1) + b_1 x_3^{**}(x_1)$, the consistent conjecture is

$$r_1^C = \frac{dG_1^{**}}{dx_1} = 1 - b_1 \frac{\alpha_2 b_2((1-\alpha_3)r_3 + \alpha_3) + ((1-\alpha_2)r_2 + \alpha_2)\alpha_3 b_3 - 2\alpha_2 b_2 \alpha_3 b_3}{((1-\alpha_2)r_2 + \alpha_2)((1-\alpha_3)r_3 + \alpha_3) - \alpha_2 b_2 \alpha_3 b_3}$$

This consistent conjecture is the unique candidate to be evolutionarily stable. Note that the consistent conjecture is less than unity, meaning that player 1 (correctly) expects the other players to partially offset an increase in his or her contribution to the public good. This exacerbates the inefficiency of the private provision of the good.

A CVE of the game is characterized by the equations

$$\alpha_1 G_1 + (1-\alpha_1)x_1 r_1 = (1-\alpha_1)m_1 r_1$$
$$\alpha_2 G_2 + (1-\alpha_2)x_2 r_2 = (1-\alpha_2)m_2 r_2$$
$$\alpha_3 G_3 + (1-\alpha_3)x_3 r_3 = (1-\alpha_3)m_3 r_3$$
$$G_1 - x_1 - b_1 x_2 - b_1 x_3 = 0$$
$$G_2 - b_2 x_1 - x_2 - b_2 x_3 = 0$$
$$G_3 - b_3 x_1 - b_3 x_2 - x_3 = 0.$$

Substituting the last three equations into the first three, the system becomes

$$((1-\alpha_1)r_1 + \alpha_1)x_1 + \alpha_2 b_2 x_2 + \alpha_3 b_3 x_3 = (1-\alpha_1)m_1 r_1$$
$$\alpha_2 b_2 x_1 + ((1-\alpha_2)r_2 + \alpha_2)x_2 + \alpha_3 b_3 x_3 = (1-\alpha_2)m_2 r_2$$
$$\alpha_3 b_3 x_1 + \alpha_2 b_2 x_2 + ((1-\alpha_3)r_3 + \alpha_3)x_3 = (1-\alpha_3)m_3 r_3.$$

Let

$$\begin{aligned}|M| = &((1-\alpha_1)r_1+\alpha_1)((1-\alpha_2)r_2+\alpha_2)((1-\alpha_3)r_3+\alpha_3)+2\alpha_1 b_1\alpha_2 b_2\alpha_3 b_3\\ &-\alpha_1 b_1((1-\alpha_2)r_2+\alpha_2)\alpha_3 b_3-((1-\alpha_1)r_1+\alpha_1)\alpha_2 b_2\alpha_3 b_3\\ &-\alpha_1 b_1\alpha_2 b_2((1-\alpha_3)r_3+\alpha_3),\end{aligned}$$

$$\begin{aligned}|M_1| = &m_1(1-\alpha_1)r_1((1-\alpha_2)r_2+\alpha_2)((1-\alpha_3)r_3+\alpha_3)+\alpha_1 b_1 m_2(1-\alpha_2)r_2\alpha_3 b_3\\ &+\alpha_1 b_1 m_2(1-\alpha_2)r_2\alpha_3 b_3-\alpha_1 b_1((1-\alpha_2)r_2+\alpha_2)m_3(1-\alpha_3)r_3\\ &-m_1(1-\alpha_1)r_1\alpha_2 b_2\alpha_3 b_3-\alpha_1 b_1 m_2(1-\alpha_2)r_2((1-\alpha_3)r_3+\alpha_3),\end{aligned}$$

$$\begin{aligned}|M_2| = &((1-\alpha_1)r_1+\alpha_1)m_2(1-\alpha_2)r_2((1-\alpha_3)r_3+\alpha_3)+m_1(1-\alpha_1)r_1\alpha_2 b_2\alpha_3 b_3\\ &+\alpha_1 b_1\alpha_2 b_2 m_3(1-\alpha_3)r_3-\alpha_1 b_1 m_2(1-\alpha_2)r_2\alpha_2 b_2-((1-\alpha_1)r_1\\ &+\alpha_1)\alpha_2 b_2 m_3(1-\alpha_3)r_3-m_1(1-\alpha_1)r_1\alpha_2 b_2((1-\alpha_3)r_3+\alpha_3),\end{aligned}$$

$$\begin{aligned}|M_3| = &((1-\alpha_1)r_1+\alpha_1)((1-\alpha_2)r_2+\alpha_2)m_3(1-\alpha_3)r_3+m_1(1-\alpha_1)r_1\alpha_2 b_2\alpha_3 b_3\\ &+\alpha_1 b_1 m_2(1-\alpha_2)r_2\alpha_3 b_3-m_1(1-\alpha_1)r_1((1-\alpha_2)r_2+\alpha_2)\alpha_3 b_3\\ &-((1-\alpha_1)r_1+\alpha_1)m_2(1-\alpha_2)r_2\alpha_3 b_3-\alpha_1 b_1\alpha_2 b_2 m_3(1-\alpha_3)r_3.\end{aligned}$$

Then $x_1^* = \frac{|M_1|}{|M|}, x_2^* = \frac{|M_2|}{|M|}, x_3^* = \frac{|M_3|}{|M|}$ and $G_1^* = \frac{|M_1|}{|M|} + b_1\frac{|M_2|+|M_3|}{|M|}$.

Evolutionarily stable conjectures of player 1 are found from the problem

$$\max_{r_1}(m_1 - x_1^*(r_1,\mathbf{r}_{-1}))^{\alpha_i}G_1^*(r_1,\mathbf{r}_{-1})^{1-\alpha_i}$$

The first-order condition for maximization is

$$(m_1-x_1^*)^{\alpha_1-1}(G_1^*)^{-\alpha_1}\left(-\alpha_1 G_1^*\frac{dx_1^*}{dr_1}+(1-\alpha_1)(m_1-x_1^*)\frac{dG_1^*}{dr_1}\right)=0.$$

Since in a CVE $-\alpha_i G_i + (1-\alpha_i)(m_i - x_i)r_i = 0$, the condition simplifies to

$$(m_1-x_1^*)^{\alpha_1}(G_1^*)^{-\alpha_1}(1-\alpha_1)\left(-r_1\frac{dx_1^*}{dr_1}+\frac{dG_1^*}{dr_1}\right)=0. \tag{6}$$

Consider $\frac{dx_1^*}{dr_1} = \frac{1}{|M|}\left(\frac{\partial|M_1|}{\partial r_1}|M| - |M_1|\frac{\partial|M|}{\partial r_1}\right)$. Since $\frac{\partial|M_1|}{\partial r_1} = m_1(1-\alpha_1)((1-\alpha_2)r_2+\alpha_2)((1-\alpha_3)r_3+\alpha_3) - m_1(1-\alpha_1)\alpha_2 b_2\alpha_3 b_3$ and $\frac{\partial|M|}{\partial r_1} = (1-\alpha_1)((1-\alpha_2)r_2+\alpha_2)((1-\alpha_3)r_3+\alpha_3) - (1-\alpha_1)\alpha_2 b_2\alpha_3 b_3$,

$$\frac{dx_1^*}{dr_1} = \frac{|K_1|}{|M|}(((1-\alpha_2)r_2+\alpha_2)((1-\alpha_2)r_3+\alpha_3)-\alpha_2 b_2\alpha_3 b_3),$$

where $|K_1| = \alpha_1(1-\alpha_1)m_1((1-\alpha_2)r_2(1-\alpha_3)r_3 + (1-\alpha_2)r_2\alpha_3(1-b_1b_3) + \alpha_2(1-\alpha_3)r_3(1-b_1b_2) + \alpha_2\alpha_3(1-b_2b_3-b_1b_3-b_1b_2+2b_1b_2b_3) + b_1(((1-\alpha_2)r_2 + \alpha_2(1-b_2))m_3(1-\alpha_3)r_3 + m_2(1-\alpha_2)r_2((1-\alpha_3)r_3 + \alpha_3(1-b_3)))) > 0$.

Consider now $\frac{dG_1^*}{dr_1} = \frac{dx_1^*}{dr_1} + b_1\left(\frac{dx_2^*}{dr_1} + \frac{dx_3^*}{dr_1}\right)$. Since $\frac{dx_2^*}{dr_1} = \frac{1}{|M|}\left(\frac{\partial |M_2|}{\partial r_1}|M| - |M_2|\frac{\partial |M|}{\partial r_1}\right)$ and $\frac{\partial |M_2|}{\partial r_1} = (1-\alpha_1)m_2(1-\alpha_2)r_2((1-\alpha_3)r_3+\alpha_3)+m_1(1-\alpha_1)\alpha_2b_2\alpha_3b_3 - (1-\alpha_1)\alpha_2b_2m_3(1-\alpha_3)r_3 - m_1(1-\alpha_1)\alpha_2b_2((1-\alpha_3)r_3+\alpha_3)$,

$$\frac{dx_2^*}{dr_1} = \frac{|K_1|}{|M|}\alpha_2b_2((1-\alpha_3r_3)+\alpha_3-a_3b_3).$$

Analogously,

$$\frac{dx_3^*}{dr_1} = \frac{|K_1|}{|M|}\alpha_3b_3((1-\alpha_2r_2)+\alpha_2-a_2b_2).$$

Therefore, $-r_1\frac{dx_1^*}{dr_1} + \frac{dG_1^*}{dr_1} = 0$ is equivalent to

$$(1-r_1)(((1-\alpha_2)r_2+\alpha_2)((1-\alpha_2)r_3+\alpha_3)-\alpha_2b_2\alpha_3b_3) +$$
$$b_1(\alpha_2b_2((1-\alpha_3r_3)+\alpha_3-\alpha_3b_3)+a_3b_3((1-\alpha_2r_2)+\alpha_2-a_2b_2)) = 0$$

and thus a candidate evolutionarily stable conjecture is

$$r_1^{ES} = 1 - b_1\frac{(a_2b_2((1-\alpha_3r_3)+\alpha_3-a_3b_3)+a_3b_3((1-\alpha_2r_2)+\alpha_2-a_2b_2))}{(((1-\alpha_2)r_2+\alpha_2)((1-\alpha_2)r_3+\alpha_3)-\alpha_2b_2\alpha_3b_3)},$$

the same as the consistent conjecture.

Now note that the left-hand side of the first order condition (6) is positive for $r_1 < r_1^{ES}$ and negative for $r_1 > r_1^{ES}$. Thus r_1^{ES} is indeed evolutionarily stable.

Proposition 2 *If the parameters of the semi-public good game of this section are such that for given r_j, r_k and consistent*

$$r_i^C = 1 - b_i\frac{(a_jb_j((1-\alpha_kr_k)+\alpha_k-a_kb_k)+a_kb_k((1-\alpha_jr_j)+\alpha_j-a_jb_j))}{(((1-\alpha_j)r_j+\alpha_j)((1-\alpha_j)r_k+\alpha_k)-\alpha_jb_j\alpha_kb_k)}$$

the CVE $(\mathbf{x}^*(\mathbf{r}), \mathbf{A}^*(\mathbf{r}))$ is interior, then conjecture r_i^C is evolutionarily stable for player i.

To illustrate the proposition, consider first the symmetric case $m_1 = m_2 = m_3 = m$, $a_1 = a_2 = a_3 = a$, $b_1 = b_2 = b_3 = b$. It is then natural to expect conjectures to be symmetric too, $r_1 = r_2 = r_3 = r$. The (mutual) consistency condition then reduces to $r = 1 - \frac{2\alpha b^2(\alpha+(1-\alpha)r-\alpha b)}{(\alpha+(1-\alpha)r)^2-\alpha^2b^2} = 1 - \frac{2\alpha b^2}{\alpha+(1-\alpha)r+\alpha b}$. This holds if $(1-\alpha)r^2 + (2\alpha+\alpha b-1)r + \alpha(2b^2-b-1) = 0$. For $r = 0$, the left-hand side is $\alpha(2b^2-b-1) < 0$ for $0 < b < 1$; for $r = 1$, the left-hand side is $2\alpha b^2 > 0$. For positive values of

Table 1 (Mutually) consistent conjectures in the public good game with $m = 1, b = 0.5$

α_1	α_2	α_3	$r_1^C = r_1^{ES}$	$r_2^C = r_2^{ES}$	$r_3^C = r_3^{ES}$	x_1^*	x_2^*	x_3^*
0.25	0.25	0.25	0.879	0.879	0.879	0.569	0.569	0.569
0.5	0.5	0.5	0.781	0.781	0.781	0.281	0.281	0.281
0.75	0.75	0.75	0.712	0.712	0.712	0.106	0.106	0.106
0.25	0.5	0.5	0.784	0.821	0.821	0.637	0.217	0.217
0.4	0.5	0.5	0.782	0.797	0.797	0.423	0.255	0.255
0.65	0.5	0.5	0.779	0.758	0.758	0.070	0.320	0.320

conjectures there is thus one consistent $r^C \in (0, 1)$, confirming the result in Costrell (1991) that consistent conjectures correspond to negative reactions, i.e. if a player increases his or her contributions, the other players decrease theirs.[14]

The proposition can be used to find consistent and evolutionarily stable conjectures also for cases that are asymmetric either in parameters or conjectures. For example, consider symmetric values of parameters $m = 1$, $b = 0.5$ and $\alpha = 0.5$. If players 2 and 3 have the (Nash) conjectures $r_2 = r_3 = 1$, then the consistent conjecture for player 1 is $r_1^C = 0.8$ (and it is evolutionarily stable because the CVE for these conjectures is interior). Table 1 shows the numerical calculations to find (mutually) consistent conjectures for some particular values of the parameters, and it also shows that the CVEs for these conjectures are interior. Therefore the consistent conjectures in Table 1 are also evolutionarily stable. Note that as the parameter α increases, less weight is put in the utility function on the public good; mutually consistent conjectures then decrease and so do contributions to the public good. The last line in the table shows that asymmetries between players should not be too large for an interior solution to exist; if the parameter α_1 increased further, x_1^* becomes 0 and the propositions cease to apply.

4.2 Contests

Consider rent-seeking contests introduced in Tullock (1980) and investigated using the techniques for aggregative games in Cornes and Hartley (2003, 2005). Each player i contributes a costly effort $x_i \geq 0$ and can win a prize of value V with probability $\frac{x_i}{x_1+\ldots+x_n}$. Each player i's payoff function is thus given by

$$u_i(x_i, A) = \frac{x_i}{A}V - c_i x_i,$$

[14]Note that if $b = 1$, then $r = 0$ is the solution of the consistency condition (Sugden 1985). However, for $r = 0$ the solution of the players' maximization problem is not interior and the first-order conditions do not characterize it. The propositions do not apply in this case.

where $A = x_1 + \ldots + x_n$ is the aggregate.[15] This aggregate is the same for all players; the game is truly aggregative. The game can still be asymmetric though, represented by possibly different marginal costs c_i of the players.

Consider player i with conjecture $r_i \geq 0$. The first-order conditions for player i's payoff maximization problem is $F_i = \frac{b_i A - b_i x_i r_i}{A^2} V - c_i = 0$, or $\frac{1}{A^2}(b_i V(A - x_i r_i) - c_i A^2) = 0$. Writing $A = x_i + A_{-i}$, the first order condition becomes $\frac{1}{A^2}(-c_i x_i^2 + ((1 - r_i)V - 2c_i A_{-i})x_i + (V - c_i A_{-i})A_{-i} = 0$. The left-hand side is negative as $x_i \to \infty$ and positive at $x_i = 0$ if $V - c_i A_{-i} > 0$. In this case, the equation

$$V(A - x_i r_i) - c_i A^2 = 0 \tag{7}$$

characterizes the choice of player i.

Consider again for illustration the case of three players ($n = 3$). The system describing a CVE is

$$\begin{aligned} V(A - x_1 r_1) - c_1 A^2 &= 0 \\ V(A - x_2 r_2) - c_2 A^2 &= 0 \\ V(A - x_3 r_3) - c_3 A^2 &= 0 \\ A - x_1 - x_2 - x_3 &= 0 \end{aligned}$$

(since there is only one aggregate, there is only one additional accounting equation). To solve the system, from the first three equations $x_i = \frac{A}{r_i V}(V - c_i A)$. Therefore $A - \frac{A}{r_1 V}(V - c_1 A) - \frac{A}{r_2 V}(V - c_2 A) - \frac{A}{r_3 V}(V - c_3 A) = 0$, or $r_1 r_2 r_3 V^3 - r_2 r_3 V^2(V - c_1 A) - r_1 r_2 V^2(V - c_2 A) - r_1 r_2 V^2(V - c_3 A) = 0$. Thus

$$A^* = \frac{r_2 r_3 + r_1 r_3 + r_1 r_2 - r_1 r_2 r_3}{r_2 r_3 c_1 + r_1 r_3 c_2 + r_1 r_2 c_3} V. \tag{8}$$

To find consistent conjectures of player 1, consider the system

$$\begin{aligned} V(A - x_2 r_2) - c_2 A^2 &= 0 \\ V(A - x_3 r_3) - c_3 A^2 &= 0 \\ A - x_1 - x_2 - x_3 &= 0. \end{aligned}$$

Solving for x_2 and x_3 from the first two equations and substituting into the third one gives $A - x_1 - \frac{A}{r_2 V}(V - c_2 A) - \frac{A}{r_3 V}(V - c_3 A) = 0$, or $(r_3 c_2 + r_2 c_3)A^2 + (r_2 r_3 - r_2 - r_3)VA - r_2 r_3 V x_1 = 0$. Using the implicit function theorem,

$$\frac{dA^{**}}{dx_1} = \frac{r_2 r_3 V}{2(r_3 c_2 + r_2 c_3)A^{**} + (r_2 r_3 - r_2 - r_3)V}.$$

[15] To avoid indeterminacies, let $u_i = \frac{1}{n}$ if $A = 0$.

Since $A^{**}(x_1^*) = A^*$, $\frac{dA^{**}}{dx_1}(x_1^*)$ can be found using A^*. Rearranging, the consistent conjecture satisfies

$$r_1 = \frac{r_2r_3(r_2r_3c_1 + r_1r_3c_2 + r_1r_2c_3)}{2(r_3c_2 + r_2c_3)(r_2r_3 + r_1r_3 + r_1r_2 - r_1r_2r_3) + (r_2r_3 - r_2 - r_3)(r_2r_3c_1 + r_1r_3c_2 + r_1r_2c_3)}.$$

Simplifying this expression leads to $(c_2r_1r_3 - c_1r_2r_3 + c_3r_1r_2)(r_2r_3 + r_1r_3 + r_1r_2 - r_1r_2r_3) = 0$. If $r_2r_3 + r_1r_3 + r_1r_2 - r_1r_2r_3 = 0$, then $A^* = 0$ thus $x_1^* = x_2^* = x_3^* = 0$, which is not an interior equilibrium. Thus consider $c_2r_1r_3 - c_1r_2r_3 + c_3r_1r_2 = 0$. The consistent conjecture of player 1 possibly leading to an interior CVE is thus

$$r_1^C = \frac{c_1r_2r_3}{r_3c_2 + r_2c_3}.$$

For evolutionary stability analysis of conjectures of player 1, consider

$$\max_{r_1} \frac{x_1^*(r_1, \mathbf{r}_{-1})}{A^*(r_1, \mathbf{r}_{-1})} V - c_1 x_1^*(r_1, \mathbf{r}_{-1}).$$

The first order condition for maximization is $\frac{1}{(A^*)^2} V \left(\frac{dx_1^*}{dr_1} A^* - x_1^* \frac{dA^*}{dr_1} \right) - c_1 \frac{dx_1^*}{dr_1} = 0$. Using Eq. (7), the condition can be rewritten as

$$\frac{Vx_1^*}{(A^*)^2} \left(r_1 \frac{dx_1^*}{dr_1} - \frac{dA^*}{dr_1} \right) = 0. \tag{9}$$

Equation (7) also implies that $x_1^* = \frac{1}{r_1} A^* - \frac{c_1}{r_1 V} (A^*)^2$. Therefore $\frac{dx_1^*}{dr_1} = \frac{1}{r_1^2} \left(\frac{dA^*}{dr_1} r_1 - A^* \right) - \frac{c_1}{V} \frac{1}{r_1^2} \left(2A^* \frac{dA^*}{dr_1} r_1 - (A^*)^2 \right)$. Equation (9) can then be written as

$$\frac{x_1^*}{(A^*)^2 r_1} \left(c_1 A^* - V - 2c_1 r_1 \frac{dA^*}{dr_1} \right) = 0.$$

Using A^* from Eq. (8), finding $\frac{dA^*}{dr_1} = \frac{-r_2r_3(c_2r_3 + c_3r_2 - c_1r_3 - c_1r_2 + c_1r_2r_3)}{(r_2r_3c_1 + r_1r_3c_2 + r_1r_2c_3)^2} V$, and substituting, the first-order condition (9) becomes

$$\frac{x_1^*}{(A^*)^2} \frac{-(c_3r_1r_2 - c_1r_2r_3 + c_2r_1r_3)(c_2r_3 + c_3r_2 - c_1r_3 - c_1r_2 + c_1r_2r_3)}{(r_2r_3c_1 + r_1r_3c_2 + r_1r_2c_3)^2} V = 0. \tag{10}$$

Therefore, provided that the CVE is interior, the first order condition is satisfied only if $c_3r_1r_2 - c_1r_2r_3 + c_2r_1r_3 = 0$, i.e.

$$r_1^{ES} = \frac{c_1r_2r_3}{c_3r_2 + c_2r_3}.$$

The unique candidate for the evolutionarily stable conjecture of player 1 is the consistent conjecture of the player.

Suppose that $c_2r_3+c_3r_2-c_1r_3-c_1r_2+c_1r_2r_3 = (c_2-c_1)r_3+(c_3-c_1)r_2+c_1r_2r_3 > 0$, which is the case unless all $r_i = 0$ or unless player 1 has marginal cost much higher than those of the other players. Then the left-hand side of (10) is positive if $r_1 < r_1^{ES}$ and negative if $r_1 > r_1^{ES}$. The consistent conjecture is then evolutionarily stable.

Proposition 3 *If the parameters of the contest game of this section are such that for given r_j, r_k and consistent*

$$r_i^C = \frac{c_i r_j r_k}{c_k r_j + c_j r_k}$$

the CVE $(\mathbf{x}^*(\mathbf{r}), \mathbf{A}^*(\mathbf{r}))$ is interior and $(c_j - c_i)r_k + (c_k - c_i)r_j + c_i r_j r_k > 0$, then conjecture r_i^C is evolutionarily stable for player i.

Consider again a few numerical examples to illustrate the proposition. Suppose that the players are symmetric, $c_1 = c_2 = c_3 = c$, and that they hold symmetric conjectures $r_1 = r_2 = r_3 = r$. Then the condition for the (mutually) consistent conjectures becomes $r = \frac{r}{2}$. The consistent conjecture is then $r^C = 0$, i.e. each player expects that a increase in his or her effort is fully offset by the decrease in the effort of the other players, leaving A unchanged. Although $A^* = V$ and $x_i^* = \frac{V}{3}$ is an interior equilibrium with such a conjecture, the proposition does not apply because the condition $(c_j - c_i)r_k + (c_k - c_i)r_j + c_i r_j r_k > 0$ is not satisfied. Indeed, if $r_j = 0$ for one of the players, then Eq. (7) becomes $A(V - A) = 0$, leading to $A = V$ in equilibrium and zero payoff to all players. Any conjecture $r_i \neq 0$ of player i implies a corner solution $x_i^* = 0$ in a CVE, again with zero payoff. Therefore $r^C = 0$ for all players is not evolutionarily stable but can be seen as *neutrally* stable for player *i*: alternative conjectures cannot lead to a higher payoff although they can be equally successful.[16]

Although the proposition does not apply to the symmetric case, it still can be used for asymmetric costs or conjectures. Table 2 shows numerical calculations for finding consistent conjectures of player 1, for given conjectures of players 2 and 3 (conjectures r_2 and r_3 are not consistent; mutually consistent conjectures are always zero for the parameters in the table). Those conjectures of player 1 are also evolutionarily stable because the conditions in Proposition 3 are satisfied. Consistent conjectures of player 1 increase with the given conjectures of the other players and with the cost of player 1 but typically stay below unity, implying that the player correctly expects the aggregate to increase by less than the increase in his or her own effort. However, it is also possible that player 1 correctly anticipates

[16]For $n = 2$, there are non-zero symmetric conjectures that are consistent and evolutionarily stable. The consistency condition for $n = 2$ is $r_i = \frac{c_i}{c_j} r_j$. If $c_i = c_j$, then any r is consistent. It is shown in Possajennikov (2009) that any $0 < r < 2$ is evolutionarily stable then.

Table 2 Consistent conjectures for player 1 in the contest game with $V = 1$

c_1	c_2	c_3	$r_1^C = r_1^{ES}$	r_2	r_3	x_1^*	x_2^*	x_3^*
1	1	1	0.375	0.75	0.75	0.406	0.203	0.203
1	1	1	0.5	1	1	0.375	0.188	0.188
1	1	1	0.625	1.25	0.125	0.344	0.172	0.172
0.75	1	1	0.281	0.75	0.75	0.925	0.027	0.027
0.75	1	1	0.375	1	1	0.764	0.076	0.076
0.75	1	1	0.469	1.25	1.25	0.655	0.010	0.010
1.25	1	1	0.781	1.25	1.25	0.200	0.194	0.194
2	1	1	1.25	1.25	1.25	0.044	0.197	0.197

that the aggregate increases by more than the increase in player's own effort if the cost and the other players' conjectures are high enough (the last line of the table). Equilibrium efforts are inversely related to cost parameters and to conjectures; but it is possible (the penultimate line in the table) that a player with a higher cost makes a higher effort in equilibrium than the other players, due to this player holding lower conjectures (that also happen to be consistent).

5 Conclusion

Richard Cornes has done much work on public good games, on contests, and on games with aggregative structure in general. Some of his work also considered conjectural variations, mostly in public good games. In this paper I also consider conjectures and I use representations of games that share some properties with aggregative games. In such representations, there is a personalized aggregate for each player; I call these representations semi-aggregative.

The idea of a semi-aggregative representation is that a player forms appropriate conjectures about how the aggregate changes and how it affects the player's payoff. In a sense, the game is reduced to just two players: the player him- or herself and the aggregate opponent. Thus the dimensionality of players' conjectures is reduced and such conjectures can be analyzed.

I show that if conjectures are subject to evolution, then only consistent conjectures can be evolutionarily stable. The result provides foundations for the (much discussed) notion of consistent conjectures as the result of evolution. On the other hand, the result can be used to find evolutionarily stable conjectures more easily, through finding first consistent conjectures. While this observation is not new for some classes of games, the result in this paper extends it to any well-behaved game.

The result is illustrated on the examples of (impure) public good games and contests. Although finding the exact value of consistent (and evolutionarily stable) conjectures in specific asymmetric games is still a difficult task (thus only 3-player examples are considered for illustration), the point of the examples is to demonstrate that it can be done, and that often consistent conjectures are indeed evolutionarily stable. The choice of public good games and contests as the examples shows that those games, to which Richard Cornes dedicated much of his work, are still a source of useful insights.

Acknowledgements I would like to thank Dirk Rübbelke and Wolfgang Buchholz for this opportunity. I would also like to thank Alex Dickson for inviting me to participate in April 2011 in a workshop on aggregative games, which incited me to think about conjectures and aggregative games, and develop the ideas leading to this paper. I also thank Maria Montero for improving the exposition in the paper.

References

Bresnahan, T. F. (1981). Duopoly models with consistent conjectures. *American Economic Review, 71*, 934–945.

Cabral, L. M. B. (1995). Conjectural variations as a reduced form. *Economics Letters, 49*, 397–402.

Corchón, L. C. (1994). Comparative statics for aggregative games. The strong concavity case. *Mathematical Social Sciences, 28*, 151–165.

Cornes, R. (2016). Aggregative environmental games. *Environmental and Resource Economics, 63*, 339–365.

Cornes, R., & Hartley, R. (2003). Risk aversion, heterogeneity and contests. *Public Choice, 117*, 1–25.

Cornes, R., & Hartley, R. (2005). Asymmetric contests with general technologies. *Economic Theory, 26*, 923–946.

Cornes, R., & Hartley, R. (2007). Aggregative public good games. *Journal of Public Economic Theory, 9*, 201–219.

Cornes, R., & Hartley, R. (2011). *Well-behaved aggregative games*. Presented at the 2011 Workshop on Aggregative Games at Strathclyde University, Glasgow.

Cornes, R., & Hartley, R. (2012). Fully aggregative games. *Economics Letters, 116*, 631–633.

Cornes, R., & Sandler, T. (1983). On commons and tragedies. *American Economic Review, 73*, 787–792.

Cornes, R., & Sandler, T. (1984a). The theory of public goods: Non-nash behavior. *Journal of Public Economics, 23*, 367–379.

Cornes, R., & Sandler, T. (1984b). Easy riders, joint production and public goods. *Economic Journal, 94*, 580–598.

Cornes, R., & Sandler, T. (1996) *The theory of externalities, public goods and club goods* (2nd ed.). Cambridge: Cambridge University Press.

Costrell, R. M. (1991). Immiserizing growth with semi-public goods under consistent conjectures. *Journal of Public Economics, 45*, 383–389.

Dixon, H. D., & Somma, E. (2003). The evolution of consistent conjectures. *Journal of Economic Behavior & Organization, 51*, 523–536.

Dockner, E. J. (1992). Dynamic theory of conjectural variations. *Journal of Industrial Economics, 40*, 377–395.

Güth, W., & Yaari, M. (1992). An evolutionary approach to explain reciprocal behavior in a simple strategic game. In U. Witt (Ed.), *Explaining process and change – Approaches to evolutionary economics* (pp. 23–34). Ann Arbor: University of Michigan Press.

Itaya, J.-I., & Dasgupta, D. (1995). Dynamics, consistent conjectures, and heterogeneous agents in the private provision of public goods. *Public Finance/Finances Publiques, 50*, 371–389.

Itaya, J.-I., & Okamura, M. (2003). Conjectural variations and voluntary public good provision in a repeated game setting. *Journal of Public Economic Theory, 5*, 51–66.

Laitner, J. (1980). Rational duopoly equilibria. *Quarterly Journal of Economics, 95*, 641–662.
Makowski, L. (1987). Are 'rational conjectures' rational? *Journal of Industrial Economics, 36*, 35–47.
Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*(Nov 2), 15–18.
Müller, W., & Normann, H.-T. (2005). Conjectural variations and evolutionary stability: A rationale for consistency. *Journal of Institutional and Theoretical Economics, 161*, 491–502.
Perry, M. K. (1982). Oligopoly and consistent conjectural variations. *Bell Journal of Economics, 13*, 197–205.
Possajennikov, A. (2009). The evolutionary stability of constant consistent conjectures. *Journal of Economic Behavior & Organization, 72*, 21–29.
Possajennikov, A. (2015). Conjectural variations in aggregative games: An evolutionary perspective. *Mathematical Social Sciences, 77*, 55–61.
Selten, R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology, 84*, 93–101.
Sugden, R. (1985). Consistent conjectures and voluntary contributions to public goods: Why the conventional theory does not work. *Journal of Public Economics, 27*, 117–124.
Tullock, G. (1980). Efficient rent-seeking. In J. M. Buchanan, R. D. Tollison, & G. Tullock (Eds.), *Toward a theory of the rent-seeking society* (pp. 131–146). College Station: Texas A&M University Press.

Mixed-Strategy Kant-Nash Equilibrium and Private Contributions to a Public Good

Ngo Van Long

1 Introduction

The theory of private contribution to public goods occupies a venerable place in the field of public economics. Explaining the incentives for private agents to contribute to public goods has been a major theoretical challenge. While some authors model the decision process of private contributors by appealing to the concept of Nash equilibrium (e.g. Warr 1983; Kemp 1984; Bergstrom et al. 1986; Cornes et al. 2001; Kemp 2009), others point to the importance of non-Nash behavior, including in particular social norms and Kantian behavior (e.g. Laffont 1975; Johansen 1976; Cornes and Sandler 1984; Brekke et al. 2003; Roemer 2010, 2015; Buchholz 2016; Buchholz et al. 2014a,b; Grafton et al. 2016).[1] In both streams of literature, most authors typically focus on equilibrium in pure strategies and they assume that the players of the game behave in a similar way.

The purpose of this paper is to model situations where contributors to a public good belong to two distinct behavioral types, which I call Kantian and Nashian. In particular, using a simple model, I consider the implications of mixed strategy equilibria in such situations. The model provides a simple theoretical account of the observed phenomenon that the extent of private provision of public goods varies considerably across time and across countries that have similar levels of per

[1]Laffont (1975) was the first paper to model Kantian behavior in the context of public goods. Cornes and Sandler (1984) made a brief reference to Kantian ethics: "…the Pareto path corresponds to Kantian behaviour, since the 'categorical imperative, whereby each acts as they want others to act, is satisfied" (p. 377). I thank Wolfgang Buchholz for this quotation.

N.V. Long (✉)
Department of Economics, McGill University, 855 Rue Sherbrooke Ouest, Montréal, QC, Canada H3A 2T7
e-mail: ngo.long@mcgill.ca

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_6

capita income.[2] To make the theory operational, I define the concept of Kant-Nash equilibrium in mixed strategies: I suppose that in the context of a game of private contribution to a public good, there are Kantian agents (in the sense of Laffont 1975, and Roemer 2010, 2015) as well as Nashian agents, and they are not restricted to using only pure strategies. This framework allows me to investigate how the defection of a Kantian into the Nashian camp can lead to a drastic decrease in the supply of the public good.

Unlike Nashian agents, who maximize personal utilities, taking as given the actions (or strategies) of other agents, Kantian agents follow behavioral norms that have been imparted on them, through moral education or other forms of transmission of moral values. Economists and other social scientists have long recognized that a society cannot function well without social norms. The most influential founding father of economic science, Adam Smith, has argued that co-operation and mutual help are incorporated in established rules of behavior, and that "upon the tolerable observance of these duties, depend the very existence of human society, which would **crumble into nothing** if mankind were not generally impressed with a reverence for those important rules of conduct."[3] Thus, according to Adam Smith, the invisible hand of the market system cannot work without the invisible hand of social norms.

The idea that Kantian behavior is pervasive in economic life was well established in the work of Adam Smith (1790), Arrow (1973) and Sen (1977, 1993); but to make it operational, some formalisation is required. A natural modeling strategy is to suppose that Kantians act **as if** they were maximizing some objective function subject to some additional side constraints that are not purely market-based. This formalisation was done by Laffont (1975) and Roemer (2010, 2015). Laffont supposes that in choices affecting public goods (such as the quality of the environment), *each Kantian agent behaves as if she believed that other Kantian agents would take 'the same action' as hers*. Laffont points out that the meaning of 'the same action' would depends on the model and would usually mean 'the same kind of action'. However to simplify the argument Laffont works with games having identical players. His purpose is to illustrate an idea rather than to present a complete model. The issue of heterogeneity among individual Kantians is taken up by Roemer (2010) and Roemer (2015) who take up the task of formalizing the concept of 'the same kind of action.' Roemer (2010) proposes the idea of 'multiplicative Kantian equilibrium', whereby a Kantian agent supposes that if she increases (or decreases)

[2]For example, the level of cleanliness of the sidewalks in a typical suburb in Kobe, Japan, is much higher than that of a typical suburb of Montreal, Canada with the same per capita income. On a positive note, Montreal's student ghetto near McGill University has gradually become cleaner over the past two decades, suggesting that the Kantian mode of behavior has become more widespread.

[3]See Smith (1790, [2002]), Part III, Chapter V, p. 190. Smith's view was echoed in Leif Johansen (1976), who wrote that economic theory "tends to suggest that people are honest only to the extent that they have economic incentives for being so", and went on to argue that "the assumption can hardly be true in its most extreme form. No society would be viable without some norms and rules of conduct."

her current effort level by a factor $\lambda > 0$, others will do likewise. In Roemer (2015), an alternative Kantian protocol is also investigated: the 'additive Kantian equilibrium.'[4]

While Laffont (1975) and Roemer (2010, 2015) focus on pure strategies, in this paper I consider mixed strategies.[5] Also, while Laffont and Roemer assume that all agents are Kantians, I model situations where in a game of private provision of public goods, some players are Kantians while other are Nashians, and I offer two alternative definitions of the Kant-Nash equilibrium.[6] This allows me to investigate how changes in the relative share of Kantians in the population may affect the outcome of the game. For example, if a Kantian agent defects from the Kantian group and becomes a Nashian, what will happen to the probability of positive provision of a public good?

One of the striking results of this paper is that when the ratio of Kantians to Nashians falls, the size of the public good may fall more than proportionately. A numerical example shows that in an economy with n Nashians and three Kantians, the size of the public good is 3, but when there are only two Kantians, the size of the public good is less than 2: it is equal to 2 only with probability 0.64, and equal to 1 (respectively zero) with probability 0.32 (respectively 0.04). Of course there is a more encouraging interpretation of this result: starting with population of 2 Kantians and $n + 1$ Nashian, a conversion of a Nashian into the Kantian camp will increase the public good more than proportionately.

2 The Model

I consider a simple game where each player has a binary action choice. There are m individuals, where $m \geq 4$. The assumption that $m \geq 4$ allows me to consider situations where there is at least one Nashian, and to explore the implications of mutations, for example when one of three Kantians becomes a Nashian. Each individual is endowed with one unit of time, which she can either 'consume' (enjoy the leisure time) or contribute to a public good project (e.g. clean up a beach or a sidewalk). Let g_i be individual i's contribution to the public good. Because of the binary choice, we have either $g_i = 0$ or $g_i = 1$. Her consumption of leisure is

[4]Roemer (2015) also introduces the 'Kantian variation' which contains additive Kantian equilibrium and multiplicative Kantian equilibrium as special cases. The Kantian variation is reminiscent of the notion of conjectural variation, which was insightfully used by Cornes and Sandler (1984) in the analysis of non-Nash behavior in public good games. See also Buchholz et al. (2014a) and Buchholz (2016).

[5]Roemer (2010) has an example of a two-person prisoner dilemma game in which mixed strategies are allowed.

[6]The concept of Kant-Nash equilibrium in pure strategies was developed in Long (2016), and has been generalized by Grafton et al. (2016).

$x_i = 1 - g_i$. The 'size' of the public good is

$$G = \sum_{i=1}^{m} g_i$$

Each individual's utility function depends only on two arguments, her consumption of leisure, x_i, and her enjoyment of the public good, G. Given my assumption of binary choice, it is convenient to specify the following utility function

$$u_i = U_i(x_{i,} G) \text{ where } x_i \in \{0, 1\}$$

and G is any non-negative integer less than or equal to m. I assume that individual i's utility ranking of the feasible bundles (x_i, G) has the following properties

$$\begin{aligned} U_i(1, m-1) > U_i(0, m) > U_i(1, m-2) > U_i(0, m-1) > \ldots > U_i(1, 1) > \\ U_i(0, 2) > U_i(1, 0) > U_i(0, 1) > U_i(0, 0). \end{aligned} \tag{1}$$

More precisely, for any two bundles (x_i, G) and (x_i', G'), where $x_i, x_i' \in \{0, 1\}$, agent i strictly prefers (x_i, G) to (x_i', G') if and only if one of the following conditions holds: (i) $x_i + G > x_i' + G_i'$, or (ii) $x_i + G = x_i' + G'$, with $x_i > x_i'$.

Note that while I assume that all players have the same ranking of their feasible vectors of consumption of the private good and the public good, I do not assume that their cardinal utility functions are identical. In particular, I allow for the possibility that the ratio of differences in utility levels across any two pairs of consumption vectors is not the same for all individuals. Thus, for example, we may have, for individuals 1 and 2,

$$\frac{U_1(1, 1) - U_1(0, 2)}{U_1(1, 0) - U_1(0, 1)} \neq \frac{U_2(1, 1) - U_2(0, 2)}{U_2(1, 0) - U_2(0, 1)}$$

The preference ranking represented by the inequalities expressed in (1) implies that if individual i behaves in the Nashian way, he will free ride on others' contribution to the public good. Thus we have an m person Prisoner Dilemma game. The Nashian agent's dominant strategy is "Do not contribute". We denote this action by D. The action 'Contribute' is denoted by C. The following questions will be addressed: (i) if some individuals are Kantian, what would be the equilibrium outcome? (ii) if a Kantian is 'converted' into the Nashian faith, how does this affect the supply of the public good?

3 Mixed-Strategy Kant-Nash Equilibrium

I assume that the set of players, $\mathcal{M} = \{1, 2, 3, \ldots, m\}$, is the union of two disjoint sets of agents denoted by $\mathcal{N}$ and $\mathcal{K}$, where

$$\mathcal{N} = \{1, 2, \ldots, n\}$$

and

$$\mathcal{K} = \{n + 1, n + 2, \ldots, n + k\}$$

Members of the sets $\mathcal{N}$ and $\mathcal{K}$ are called Nashian agents and Kantian agents respectively. Roughly speaking, when a Nashian makes a decision on whether he should stick to his current equilibrium action or to deviate from it, he considers the effect of that deviation on his utility, on the assumption that all other players stick to their current actions. In other words, a Nashian agent's "counterfactual" is the scenario in which he deviates while others do not. In contrast, for any Kantian agent $j \in \mathcal{K}$, I define her counterfactual as the scenario in which if she deviates from her equilibrium action, all other members of set $\mathcal{K}$ will deviate likewise.

More formally, denote by p_i the probability that player i chooses C (and by $1-p_i$ the probability that she chooses D). The vector $\mathbf{p}$ denotes the mixed strategy profile $(p_1, p_2, \ldots, p_m)$. Since there are n Nashian players and k Kantian players, I will also use the notation

$$\mathbf{p} = \left(\mathbf{p}^N, \mathbf{p}^K\right) = (p_1^N, p_2^N, \ldots, p_n^N, p_{n+1}^K, \ldots, p_{n+k}^K).$$

Consider a candidate equilibrium vector $\mathbf{p}$. I denote by $V_i(\mathbf{p})$, or, equivalently, $V_i(\mathbf{p}_{-i}, p_i)$ the expected utility of player i. A Nashian agent $i \in \mathcal{N}$ is satisfied with his strategy p_i if and only if there is no alternative strategy $p_i' \neq p_i$ that yields him a higher expected utility. Concerning Kantian agents, I posit that when a Kantian agent $n + j \in \mathcal{K}$ considers modifying her p_{n+j} by any amount δ (such that $p_{n+j} + \delta$ is bounded above by 1 and bounded below by 0), she evaluates the consequence of her action on her utility, using the counterfactual that all other Kantian players $n + s \in \mathcal{K}$ would modify their mixed strategies in a 'similar' way. Here, I take 'similar' to mean that p_{n+s} will become $p'_{n+s} = p_{n+s} + \delta$, with the exception that when $p_{n+s} + \delta$ is outside the unit interval [0, 1], the Kantian agent $n+j$ will suppose, instead, that $p'_{n+s} = 1$ (if $p_{n+s} + \delta > 1$) or $p'_{n+s} = 0$ (if $p_{n+s} + \delta < 0$). Another way of putting this is that, when the Kantian agent $n + j$ contemplates deviation by δ, she supposes that other Kantian agents $n + s$ would adjust their probability p_{n+s} by the following amount

$$\phi_{n+s}(\delta) \equiv \min\left\{\min\left[|\delta|, 1 - p_{n+s}\right], \max\left[\delta, -p_{n+s}\right]\right\}.$$

This means that if $\delta \geq 0$, then $\phi_{n+s}(\delta) = \min[|\delta|, 1 - p_{n+s}]$, and if $\delta < 0$, then $\phi_{n+s}(\delta) = \max[\delta, -p_{n+s}]$.

Formally, I define a mixed-strategy Kant-Nash equilibrium as follows:

Definition 1 (Mixed Strategy Kant-Nash Equilibrium) A mixed strategy profile $\mathbf{p} = (\mathbf{p}^N, \mathbf{p}^K)$ is a Kant-Nash equilibrium if and only if

(i) for all $i \in \mathcal{N}$,

$$V_i(\mathbf{p}_{-i}, p_i) \geq V_i(\mathbf{p}_{-i}, p_i') \text{ for all } p_i' \in [0, 1]$$

and

(ii) for all $n + j \in \mathcal{K}$,

$$V_{n+j}(\mathbf{p}^N, \mathbf{p}^K) \geq V_{n+j}(\mathbf{p}^N, p_{n+1}^K + \phi_{n+1}(\delta), \ldots, p_{n+j}^K + \delta, .., p_{n+k}^K + \phi_{n+k}(\delta))$$

for all $\delta \in \left[-p_{n+j}^K, 1 - p_{n+j}^K\right]$, where

$$\phi_{n+s}(\delta) \equiv \min\left\{\min\left[|\delta|, 1 - p_{n+s}\right], \max\left[\delta, -p_{n+s}\right]\right\}.$$

4 Analysis: Mixed Strategy Equilibrium

In this section, I consider mixed strategy Kant-Nash equilibria in the game of private contribution to the public good. Clearly, since D is the dominant strategy for all Nashian players, we have $p_i^N = 0$ for all $i \in \mathcal{N}$. That is, the Nashian players do not contribute to the public good. Since there is perfect information, the k Kantian players know that the size of the public good cannot be greater than k. Can I find a vector $(p_{n+1}^K, p_{n+2}^K, \ldots, p_{n+k}^K)$ such that $(0, 0, .., 0, p_{n+1}^K, p_{n+2}^K, \ldots, p_{n+k}^K)$ is a Kant-Nash equilibrium?

Let me simplify the exposition by focusing on the case where the number of Kantians is $k = 3$, and the number of Nashians is any positive integer n. Since we know that the Nashians will not contribute to the public good, it suffices to consider the actions of the Kantians. Denote the three Kantian agents by α, β and ω.With only three Kantian agents, I can depict the various outcomes using a simple matrix representation. For the moment, consider pure strategies. If agent ω chooses the pure strategy "Contribute" (i.e., $p_\omega = 1$), then, depending on which pure strategy that players α and β use, the payoffs to the three agents are reported in the four boxes of matrix A below. For example, given $p_\omega = 1$, if both α and β contribute with probability 1, the payoffs of the three agents are $U_\alpha(0, 3)$, $U_\beta(0, 3)$ and $U_\omega(0, 3)$.

	Matrix A ($p_\omega = 1$)	
	C_β	D_β
C_α	$U_\alpha(0,3)$ $U_\beta(0,3)$ $U_\omega(0,3)$	$U_\alpha(0,2)$ $U_\beta(1,2)$ $U_\omega(0,2)$
D_α	$U_\alpha(1,2)$ $U_\beta(1,2)$ $U_\omega(0,2)$	$U_\alpha(1,1)$ $U_\beta(1,1)$ $U_\omega(0,1)$

If agent ω chooses the pure strategy "Do not contribute" (i.e., $p_\omega = 0$), then, depending on which pure strategy players α and β use, the payoffs to the three agents are displayed in matrix B below.

	Matrix B ($p_\omega = 0$)	
	C_β	D_β
C_α	$U_\alpha(0,2)$ $U_\beta(0,2)$ $U_\omega(1,2)$	$U_\alpha(0,1)$ $U_\beta(1,1)$ $U_\omega(1,1)$
D_α	$U_\alpha(1,1)$ $U_\beta(0,1)$ $U_\omega(1,1)$	$U_\alpha(1,0)$ $U_\beta(1,0)$ $U_\omega(1,0)$

When agents α and β make their decisions on p_a and p_β, they do not know what strategy is chosen by agent ω. Their expected payoffs if $p_\omega = 1$ can be computed using the entries in matrix A :

$$\begin{aligned} V_\alpha(p_\alpha, p_\beta \,|p_\omega = 1) = {} & U_\alpha(0,3)p_\alpha p_\beta + U_\alpha(1,2)(1-p_\alpha)p_\beta \\ & + U_\alpha(0,2)p_\alpha(1-p_\beta) + U_\alpha(1,1)(1-p_\alpha)(1-p_\beta) \end{aligned} \tag{2}$$

$$\begin{aligned} V_\beta(p_\alpha, p_\beta \,|p_\omega = 1) = {} & U_\beta(0,3)p_\alpha p_\beta + U_\beta(1,2)(1-p_\beta)p_\alpha \\ & + U_\beta(0,2)p_\beta(1-p_\alpha) + U_\beta(1,1)(1-p_\alpha)(1-p_\beta) \end{aligned} \tag{3}$$

Similarly, their expected payoffs if $p_\omega = 0$ can be computed using the entries in matrix B :

$$\begin{aligned} V_\alpha(p_\alpha, p_\beta \,|p_\omega = 0) = {} & U_\alpha(0,2)p_\alpha p_\beta + U_\alpha(1,1)(1-p_\alpha)p_\beta \\ & + U_\alpha(0,1)p_\alpha(1-p_\beta) + U_\alpha(1,0)(1-p_\alpha)(1-p_\beta) \end{aligned} \tag{4}$$

$$\begin{aligned} V_\beta(p_\alpha, p_\beta \,|p_\omega = 0) = {} & U_\beta(0,2)p_\alpha p_\beta + U_\beta(1,1)(1-p_\beta)p_\alpha \\ & + U_\beta(0,1)p_\beta(1-p_\alpha) + U_\beta(1,0)(1-p_\alpha)(1-p_\beta) \end{aligned} \tag{5}$$

Thus, player α's expected payoff under the Kantians' mixed strategy profile $(p_\alpha, p_\beta, p_\omega)$ is

$$V_\alpha(p_\alpha, p_\beta, p_\omega) = p_\omega V_\alpha(p_\alpha, p_\beta \,|p_\omega = 1) + (1-p_\omega)V_\alpha(p_\alpha, p_\beta \,|p_\omega = 0) \tag{6}$$

To simplify notation, I normalize by setting $U_\alpha(0,3) = 1$ and $U_\alpha(1,0) = 0$, and let $U_\alpha(1,2) = a_\alpha$, $U_\alpha(1,1) = b_\alpha$, $U_\alpha(0,2) = f_\alpha$ and $U_\alpha(0,1) = g_\alpha$. (This normalisation does not affect any result.) Then, because of the ranking given by (1) above, we must have

$$a_\alpha > 1 > b_\alpha > f_\alpha > 0 > g_\alpha \tag{7}$$

and using Eqs. (2), (4) and (6), we obtain

$$\begin{aligned} V_\alpha = {} & p_\alpha p_\beta p_\omega + a_\alpha(1-p_\alpha)p_\beta p_\omega + f_\alpha \left[p_\alpha(1-p_\beta)p_\omega + p_\alpha p_\beta(1-p_\omega)\right] \\ & + b_\alpha \left[(1-p_\alpha)(1-p_\beta)p_\omega + (1-p_\alpha)p_\beta(1-p_\omega)\right] + g_\alpha p_\alpha(1-p_\beta)(1-p_\omega) \end{aligned} \tag{8}$$

Similar expressions can be obtained for V_β and V_ω:

$$\begin{aligned} V_\beta = {} & p_\alpha p_\beta p_\omega + a_\beta(1-p_\beta)p_\alpha p_\omega + f_\beta \left[p_\beta(1-p_\alpha)p_\omega + p_\alpha p_\beta(1-p_\omega)\right] \\ & + b_\beta \left[(1-p_\alpha)(1-p_\beta)p_\omega + (1-p_\beta)p_\alpha(1-p_\omega)\right] + g_\beta p_\beta(1-p_\alpha)(1-p_\omega) \end{aligned} \tag{9}$$

$$\begin{aligned} V_\omega = {} & p_\alpha p_\beta p_\omega + a_\omega(1-p_\omega)p_\alpha p_\beta + f_\omega \left[p_\omega(1-p_\beta)p_\alpha + p_\omega p_\beta(1-p_\alpha)\right] \\ & + b_\omega \left[(1-p_\omega)(1-p_\beta)p_\alpha + (1-p_\beta)p_\omega(1-p_\alpha)\right] + g_\omega p_\beta(1-p_\omega)(1-p_\alpha) \end{aligned} \tag{10}$$

We can then proceed to compute a Kant-Nash equilibrium in mixed strategies, and show how the equilibrium strategies of the three Kantian players depend on (a_j, b_j, f_j, g_j), where $j = \alpha, \beta, \omega$. (The Nashian players always choose the pure strategy D, i.e., $p_i^N = 0$.) For simplicity, I will focus on the case where the three Kantian players have the same cardinal utility function, i.e.

$$(a_j, b_j, f_j, g_j) = (a, b, f, g) \text{ for } j = \alpha, \beta, \omega \tag{11}$$

where

$$a > 1 > b > f > 0 > g \tag{12}$$

Under this assumption, we can focus on the symmetric Kant-Nash equilibrium in which all three Kantians choose the same mixed strategy, $p_\alpha = p_\beta = p_\omega = p^*$. Using (8), (11) and Definition 1, it is straightforward to see that the equilibrium p^* must maximize the cubic function $V(p)$

$$V(p) = \gamma_0 p^3 + \gamma_1 p^2 + \gamma_2 p$$

subject to $0 \leq p \leq 1$, where

$$\gamma_0 = (2b + g) - (a + 2f) + 1$$

$$\gamma_1 = (a + 2f) - 2(2b + g)$$

and

$$\gamma_2 = 2b + g$$

Note that $V(0) = 0$ and $V(1) = 1$.

The following proposition can be proved.

Proposition 1 *Assume that the Kantian agents have the same cardinal utility function. Consider an economy with n Nashians and* 3 *Kantians (where n is any non-negative integer). The symmetric Kant-Nash equilibrium has the following properties:*

(i) If $a + 2f > 3$*, then at the symmetric Kant-Nash equilibrium, Kantian agents will choose a non-degenerate mixed strategy* p^**, where* $0 < p^* < 1$*:*

$$p^* = \frac{2\gamma_1 + \sqrt{4\gamma_1^2 - 12\gamma_0(2b+g)}}{-6\gamma_0} < 1 \tag{13}$$

(ii) If $a + 2f < 3$*, then at the symmetric Kant-Nash equilibrium, Kantian agents contribute with probability* 1.

Proof See the Appendix.

Example 1.1 (The Kantians Use a Non-degenerate Mixed Strategy) Let us set $a = 1.7, b = 0.9, f = 0.7, g = -0.1$, so that condition (12) is satisfied. Then $\gamma_0 = -0.4$, $\gamma_1 = -0.3$, $\gamma_2 = 1.7$, and $\Delta = 4(\gamma_1)^2 - 12(\gamma_0)(\gamma_2) = 8.52$.

The positive root is $p^* = 0.966\,21$. The Kantian agents contribute most of the times, but not always.[7]

[7] In a model with only two Kantians and no Nashian, Roemer (2010) also showed the possibility of a mixed strategy equilibrium. He wrote that "only if utility from cheating is not too high will Kantian ethics induce full cooperation." That statement is somewhat misleading, as it incorrectly equates "full cooperation" with "always contributing." It should be restated as follows: The strategy profile "always contributing" (i.e. choosing C with probability 1) is a collusive equilibrium if and only if a player's personal gain from non-contributing (while others contribute) is sufficiently low". In fact, when a player's personal gain from non-contributing (while others contribute) is high enough, all players will agree that they will be better off by choosing a common non-degenerate mixed strategy, and a social planner would advise them to do the same thing.

Example 1.2 (The Kantians Contribute with Probability 1) Let $a = 1.3$, $b = 0.9$, $f = 0.3$, $g = -0.1$. In this case, a is not sufficiently large to make mixed strategies attractive. All three Kantians will contribute with probability 1.

Proposition 1 implies that if the utility of an individual to free ride on others is sufficiently large, so that $a + 2f > 3$, the symmetric Kant-Nash equilibrium will not ensure that all Kantians contribute to the public good all the times. Upon reflection, this property does not imply a failure of Kantian ethics. In fact, when $a + 2f > 3$, if there were a social planner that seeks to maximize the expected utility of the representative Kantian, taking as given the choice D of the Nashians, she would ask each Kantian to contribute with the non-degenerate probability p^* given by (13).

5 Defection from the Kantian Camp

Let us assume that $a + 2f < 3$ holds, so that in an economy with n Nashians and 3 Kantians, in equilibrium each Kantian will contribute to the public good with probability 1. Then the size of the public good is 3. Now assume that due to random mutation, the Kantian population falls to 2 and the Nashian population increases to $n + 1$. Would the remaining two Kantians continue to contribute with probability 1?

Suppose the previously Kantian agent ω becomes Nashian. Then she will choose the dominant strategy D. What is the mixed-strategy Kant-Nash equilibrium in this new economy with just two Kantians and $n + 1$ Nashians? Matrix B in the preceding section can be used to find the symmetric Kahn-Nash equilibrium. We obtain the following proposition.

Proposition 2 *Assume that Kantian agents have the same cardinal utility function. Consider an economy with n Nashians and 3 Kantians, where n is any non-negative integer. Suppose that due to mutation, one Kantian becomes Nashian. Then the resulting Kant-Nash equilibrium has the following properties.*

(i) If $f < \frac{b+g}{2}$, then at the symmetric Kant-Nash equilibrium, the two remaining Kantian agents will choose a non-degenerate mixed strategy $p^{**} \in (0, 1)$, where

$$p^{**} = \frac{b + g}{2(b + g) - 2f} < 1$$

(ii) If $f \geq \frac{b+g}{2}$ then at the symmetric Kant-Nash equilibrium, the two remaining Kantian agents will choose the pure strategy $p = 1$ (they contribute to the public good with probability 1).

Proof See the Appendix.

Example 2.1 (Kantians Contribute with Probability 1) Let $b = 0.9, g = -0.1$ and $f = 0.7$. Then the collusive strategy for the two remaining Kantian is to contribute with probability 1.

Example 2.2 (Kantians Contribute with Probability Less than 1) Let $b = 0.9, g = -0.1$ and $f = 0.3$. Then the collusive strategy for the two remaining Kantian is to contribute with probability $p^{**} = 0.8$.

Taking into account part (ii) of Proposition 1, we can infer from part (i) of Proposition 2 that if b is sufficiently large relative to f, there exists cases where, with three Kantians, the equilibrium size of the public good is 3 (as in Example 1.2) but after the defection of one Kantian, the remaining two Kantians no longer contribute with probability 1 (Example 2.2). In that case, the post-defection expected size of the public good is strictly smaller than 2.

6 An Alternative Equilibrium Concept: The Inclusive Kant-Nash Equilibrium

In the preceding section, we used Definition 1 (mixed strategy Kant-Nash equilibrium) to analyse a game of private contribution to a public good. That definition says that a Kantian agent $n+j$ (where $j = 1, 2, \ldots, k$) would be satisfied with her current mixed strategy p^K_{n+j} only if any change in p^K_{n+j} by some $\delta \in \left[-p^K_{n+j}, 1 - p^K_{n+j}\right]$ will result in a 'new expected utility' level that exceeds her current utility, where the 'new expected utility' is computed on the two-fold supposition that

(i) all other Kantians $n + s \in \mathcal{K}$ will also change their p^K_{n+s} by $\phi_{n+s}(\delta) = \min\{\min[|\delta|, 1 - p_{n+s}], \max[\delta, -p_{n+s}]\}\, t$, and
(ii) all Nashians maintain their current strategies p^N_i, $i \in \mathcal{N}$.

Definition 1 makes sense if we posit that Kantians are 'realistic' and if we insist that expectations are self-confirming both in and out of equilibrium. However, there is an alternative view about what would be appropriate Kantian behavior, which is perhaps closer to the Kantian ethics. According to Kant's categorical imperative, 'Act as if the maxim of your action were to become through your will a general natural law', a Kantian agent should take an action (e.g. contribute to a public good) if she would *wish* that others would take similar actions under similar circumstances.[8] There is no requirement that her wish about others' behavior would be expected to be fulfilled. The appropriate action should be taken, even if she knows that Nashians would not act likewise. The Kantian ethics do not require

[8]See Bertrand Russell's *A History of Western Philosophy*. As explained by Russell (1945, p. 737) 'There are two sorts of imperative: the hypothetical imperative, which says "You must do so-and-so if you wish to achieve such-and-such an end"; and the categorical imperative, which says that a certain kind of action is objectively necessary, without regard to any end.'

realistic expectations. In this section, therefore, I propose an alternative concept of Kant-Nash equilibrium, which I call “Inclusive Kant-Nash equilibrium’, because it requires that Kantians include all agents (Kantians and Nashians) in their wish that everyone would behave the same way. Formally, in the context of the game of private contribution to a public good, I propose the following Definition:

Definition 2 (Mixed-Strategy Inclusive Kant-Nash Equilibrium) A mixed strategy profile $\mathbf{p}$ is an Inclusive Kant-Nash equilibrium if and only if

(i) for all $i \in \mathcal{N}$,

$$V_i(\mathbf{p}_{-i}, p_i) \geq V_i(\mathbf{p}_{-i}, p_i') \text{ for all } p_i' \in [0, 1]$$

and

(ii) for all $n + j \in \mathcal{K}$, where $j = 1, 2, \ldots, k$,

$$V_{n+j}(\mathbf{p}^N, \mathbf{p}^K) \geq V_{n+j}(p_1^N + \phi_1(\delta), \ldots, p_n^N + \phi_n(\delta), p_{n+1}^K + \phi_{n+1}(\delta), \ldots, p_{n+j}^K + \delta, .., p_{n+k}^K + \phi_{n+k}(\delta))$$

for all $\delta \in \left[-p_{n+j}^K, 1 - p_{n+j}^K\right]$, and where

$$\phi_v(\delta) \equiv \min\left\{\min\left[|\delta|, 1 - p_v^K\right], \max\left[\delta, -p_v^K\right]\right\}, \text{ for all } v \in \mathcal{K} \cup \mathcal{N}.$$

Requirement (ii) in Definition 2 means that each Kantian agent j is contented with what she is currently doing only if there is no feasible addition of δ to p_j^K that would increase her expected utility, assuming all other agents (Kantians and Nashians) would change their strategy in a similar way. The following proposition states an obvious relationship between the two equilibrium concepts.

Proposition 3 *In the context of the public good game with binary choice of effort level, any Kant-Nash equilibrium* $(p_1^N, p_2^N, \ldots, p_n^N, p_{n+1}^K, \ldots, p_{n+k}^K)$ *(under Definition 1) which has the property that* $p_{n+1}^K = p_{n+2}^K \cdots = p_{n+k}^K = 1$ *is also an Inclusive Kant-Nash equilibrium under Definition 2.*

Proof Consider a Kant-Nash equilibrium $(p_1^N, p_2^N, \ldots, p_n^N, p_{n+1}^K, \ldots, p_{n+k}^K)$ (under Definition 1) that has the property that $p_{n+1}^K = p_{n+2}^K \cdots = p_{n+k}^K = 1$. Clearly any positive deviation $\delta > 0$ contemplated by a Kantian agent $n + s \in \mathcal{K}$ is not a feasible choice. So we are left with the possibility of negative deviation, $\delta < 0$. Since the Nashians always choose $p_1^N = p_2^N = \ldots = p_n^N = 0$, any negative deviation $\delta < 0$ contemplated by a Kantian would imply no change in p_i^N, and therefore cannot improve the expected utility of the Kantian agent. Hence under these circumstances, condition (ii) in Definition 2 is satisfied when condition (ii) in Definition 1 is satisfied.

Proposition 3 is very special, because it applies only to the case where the Kant-Nash equilibrium (according to Definition 1) is a profile of degenerate mixed strategies. However, if a given symmetric Kant-Nash equilibrium found in the preceding section (under Definition 1) has the property that the Kantians play a non-degenerate mixed strategy, then in general that equilibrium is different from the Inclusive Kant-Nash equilibrium in the sense of Definition 2. This is demonstrated below by way of an example.

Consider for simplicity an economy with three agents: α and β are Kantians and η is Nashian. To simplify, we assume that α and β have the same cardinal utility function, i.e., $a_\alpha = a_\beta = a$, $b_\alpha = b_\beta = b$, $f_\alpha = f_\beta = f$ and $g_\alpha = g_\beta = g$. From our previous analysis (see Proposition 2) using Matrix B, we find that if $f < \frac{b+g}{2}$, then at the symmetric Kant-Nash equilibrium, the Kantian agents α and β choose mixed strategy $p^{**} \in (0, 1)$, where

$$p^{**} = \frac{b+g}{2(b+g)-2f} < 1 \tag{14}$$

We now show that in general p^{**}, as given by (14) is different from the strategy that Kantians use in an Inclusive Kant-Nash equilibrium.

Since the Nashian agent η will use his dominant strategy D, we have $p_\eta = 0$. For the triple $(p_\eta, p_\alpha, p_\beta)$, with $p_\eta = 0$, and $p_\alpha = p_\beta = p'$ to be an Inclusive Kant-Nash Equilibrium, we require that, for all $\delta \in [-p', 1-p']$, the expected utility of α and β, evaluated at $(p_\eta, p_\alpha, p_\beta) = (0 + \max(\delta, 0), p' + \delta, p' + \delta)$, is smaller than their expected utility evaluated at $(0, p_\alpha, p_\beta)$. Using Eqs. (8) and (9), $p_\alpha = p_\beta = p' + \delta$ and with p_ω replaced by the hypothetical mixed strategy of the Nashian agent η, namely $p_\eta = 0 + \max(0, \delta)$, we have

$$\begin{aligned} V_{\alpha,\beta}(p\,|p_\eta) &= p^2 p_\eta + a(1-p)pp_\eta + f\left[p(1-p)p_\eta + p^2(1-p_\eta)\right] \\ &\quad + b\left[(1-p)^2 p_\eta + (1-p)p(1-p_\eta)\right] + gp(1-p)(1-p_\eta) \end{aligned}$$

where p stands for $p' + \delta$.

For $(0, p', p')$ to be an Inclusive Kant-Nash equilibrium, we must verify that the following conditions are satisfied:

$$V_{\alpha,\beta}(p'\,|p_\eta = 0) \geq V_{\alpha,\beta}(p' + \delta\,|p_\eta = 0 + \delta) \text{ for all } \delta \in \left[0, 1-p'\right] \tag{15}$$

and

$$V_{\alpha,\beta}(p'\,|p_\eta = 0) \geq V_{\alpha,\beta}(p' + \delta\,|p_\eta = 0) \text{ for all } \delta \in \left[-p', 0\right] \tag{16}$$

Clearly, in general a non-degenerate mixed strategy profile $(0, p^*, p^*)$ where $p^* \in (0, 1)$ which satisfies the properties required by Definition 1 may not satisfy the properties required by Definition 2. It suffices to consider the following example.

Example 3.1 Let $a = 1.3$ $b = 0.9, f = 0.3, g = -0.1$. Assume there are two Kantians, α and β, and one Nashian, η. Then we know from Example 2.2 that at the symmetric Kant-Nash equilibrium, each of the two Kantians will contribute with probability 0.8, i.e., the equilibrium mixed strategy profile is $(p_\eta, p_\alpha, p_\beta) = (0, \frac{8}{10}, \frac{8}{10})$. Can this profile also be a symmetric Inclusive Kant-Nash equilibrium? The answer is no. At an Inclusive Kant-Nash equilibrium, the Kantian players compare their utility under the candidate equilibrium with the utility they would obtain if they deviate, assuming that the Nashian agent would deviate likewise. Therefore, the entries in Matrix A become relevant in their calculation, though they were not relevant under the (standard, non-inclusive) Kant-Nash equilibrium with two Kantians and one Nashian agent,

The right-hand side of Eq. (15) is

$$F(\delta\,|p') \equiv p^2\delta + a(p-p^2)\delta + f\left[(p-p^2)\delta + p^2(1-\delta)\right] \\ +b\left[(1-p)^2\delta + (p-p^2)(1-\delta)\right] + g(p-p^2)(1-\delta)$$

where $p = p' + \delta$. Then $(0, p', p')$ is a non-degenerate mixed strategy Inclusive Kant-Nash equilibrium only if $F_\delta(0\,|p') = 0$ and $F_{\delta\delta}(0\,|p') < 0$. Now

$$F_\delta(0|p') = p^2 + a(p-p^2) + f\left[(p-p^2) - p^2 + 2p\right] \\ +b\left[(1-p)^2 - (p-p^2) + (1-2p)\right] + g\left[(p-2p) - (p-p^2)\right]$$

where p is evaluated at p'. Simplify to get

$$F_\delta(0|p') = (2b - a - 2f + g + 1)\,p^2 + (a - 5b + 3f - 2g)\,p + 2b \tag{17}$$

Evaluating the above expression at $p' = 0.8$, using $a = 1.3$ $b = 0.9, f = 0.3, g = -0.1$, we get $F_\delta(0|p') = 0.312 > 0$. Thus, at $(p_\eta, p_\alpha, p_\beta) = (0, \frac{8}{10}, \frac{8}{10})$ the Kantians find that a small deviation by some small $\delta > 0$ will increase their utility, assuming that the Nashian would follow suit. (They know that this assumption is 'unrealistic', but for the purpose of making a morally correct decision according to the Inclusive Kantian criterion, that assumption has to be made.) Can $F_\delta(0|p') = 0$ at some $p' \in (0.8, 1]$? Using the specified values of the parameter, $F_\delta(0|p')$ becomes the following polynomial of degree 2 in p

$$\psi(p) \equiv (0.8)p^2 - 2p + 1.8$$

Clearly $\psi(p)$ cannot be equal to zero for any real value of p. We conclude that a symmetric Inclusive Kant-Nash equilibrium can only occur at $(p_\eta, p_\alpha, p_\beta) = (0, 1, 1)$.

7 Concluding Remarks

Using a simple model of private contribution to a public good, where some agents are motivated by the Kantian ethics and others behave in the standard Nashian way, I have shown that when the number of Kantians changes, there may be a switch from pure strategy equilibrium to mixed strategy equilibrium. A decline in the population share of Kantians may lead to a disproportionate fall in the expected size of the public good, with a positive probability of no public good provision. An interesting extension of the model would be to study mixed strategy equilibria in a dynamic game of contribution to a public good (along the lines of Cornes et al. 2001), allowing for dynamic interactions among Kantians and Nashians, using the techniques of differential games (Dockner et al. 2000).

Another direction for extension of the model would be to endogenize the ratio of Kantians to Nashians, and to model the endogenous evolution of that ratio. One possibility would be to consider some process of replicator dynamics.[9] If there are enough Kantians, a momentum may be generated to create growth in the number Kantians, and consequently outcomes that were previously unthinkable will become feasible. The following remark by Kay (2015), though written a slightly different context, is quite relevant here

> The limits of what is politically possible have changed so much and so often in the course of my lifetime . . . that to feel constrained by what is 'politically possible' is simply a failure of imagination. (Kay 2015, p. 307).

Acknowledgements I am grateful to Scott Finley, Guy Laroque, Rohan Dutta and Eytan Sheshinsky for their comments on an earlier version of this paper, and to Geir Asheim, Wolfgang Buchholz, Dirk Rübbelke, and Rüdiger Pethig for helpful discussion.

Appendix

Proof of Proposition 1

Let $p_i \in [0, 1]$ be the probability that player i chooses C, where $i = \alpha, \beta, \omega$. Let $q_i = 1 - p_i$.

Then under the mixed strategy profile $(p_\alpha, p_\beta, p_\gamma)$, the expected utility of player α is

$$\begin{aligned} V_\alpha = {} & (1 \times p_\alpha p_\beta p_\omega) + (f \times p_\alpha q_\beta p_\omega) + \left(a \times q_\alpha p_\beta p_\omega\right) + \left(b \times q_\alpha q_\beta p_\omega\right) \\ & + \left(f \times p_\alpha p_\beta q_\omega\right) + \left(g \times p_\alpha q_\beta q_\omega\right) + \left(b \times q_\alpha p_\beta q_\omega\right) \end{aligned}$$

[9]For examples of replicator dynamics, see Bala and Long (2005), Breton et al. (2010), Sethi and Somanathan (1996). For an account of the evolution of cooperation in human society, see Seabright (2010).

Consider the situation where the three players can collude. Suppose they agree on doing the same thing. Let p be their common probability of choosing action C. Then the expected payoff to each of the three players under this symmetric strategy profile is

$$\begin{aligned} V(p) &= p^3 + ap^2(1-p) + 2bp(1-p)^2 + 2fp^2(1-p) + gp(1-p)^2 \\ &= (2b - a - 2f + g + 1)\,p^3 + (a - 4b + 2f - 2g)\,p^2 + (2b + g)\,p \end{aligned}$$

$$V(p) = \gamma_0 p^3 + \gamma_1 p^2 + \gamma_2 p + \gamma_3$$

where $\gamma_3 = 0$. Clearly, $V(0) = 0$ and $V(1) = 1$. This shows that (C, C, C) yields higher utility than (D, D, D). Restricting p to the domain $[0, 1]$, we want to find out whether $V(p)$ attains its maximum at $p = 1$, or at some $0 < p < 1$.

Remark A1 If $\gamma_0 \neq 0$, then the product of the roots of quadratic equation $V'(p) = 0$ is

$$\lambda_1 \lambda_2 = \frac{2b + g}{6\gamma_0}$$

Remark A2 It is useful to consider the function $V(x)$ defined over the entire real line:

$$V(x) \equiv \gamma_0 x^3 + \gamma_1 x^2 + \gamma_2 x \text{ where } x \in (-\infty, +\infty)\,.$$

We have

$$V'(x) = 3\gamma_0 x + 2\gamma_1 x + \gamma_2$$

Thus $V'(0) = 2b + g$ and

$$V'(1) = 3\gamma_0 + 2\gamma_1 + \gamma_2 = 3 - 2f - a$$

Lemma A1

(i) $V'(1) < 0$ *is equivalent to* $a + 2f > 3$.
(ii) $V'(1) < 0$ *implies* $\gamma_0 < 0$.
(iii) If, in addition to $V'(1) < 0$ *(and thus* $\gamma_0 < 0$*) we have* $b + 2g > 0$*, then the cubic equation* $V(x) = 0$ *has a positive root, a negative root, and a zero root, and consequently, restricting p in the unit interval* $[0, 1]$*,* $V(p)$ *attains its maximum at the unique root* $p^* \in (0, 1)$ *of the quadratic equation* $V'(p) = 0$ *(the other root being negative). More precisely,*

$$p^* = \frac{2\gamma_1 + \sqrt{\Delta}}{-6\gamma_0}$$

where

$$\Delta = 4\gamma_1^2 - 12\gamma_0(b + 2g) < 4\gamma_1^2$$

(iv) Given $V'(1) < 0$ *(and thus* $\gamma_0 < 0$*), if we have* $b+2g < 0$*, then* $\gamma_1 > 0$ *and the cubic equation* $V(x) = 0$ *has a zero root, and two positive roots, of which only one is in the interior of* $[0, 1]$*. Consequently, restricting p in the unit interval* $[0, 1]$*,* $V(p)$ *attains its maximum at a unique* $p^* \in (0, 1)$*, where* $p^\#$ *is the bigger root of the quadratic equation* $V'(p) = 0$*. Specifically,*

$$p^* = \frac{2\gamma_1 + \sqrt{\Delta}}{-6\gamma_0}$$

with

$$\Delta = 4\gamma_1^2 - 12\gamma_0(b + 2g) > 4\gamma_1^2$$

and $V(p)$ *attains its minimum at* $\widetilde{p} \in \left(0, p^\#\right)$ *where*

$$\widetilde{p} = \frac{2\gamma_1 - \sqrt{\Delta}}{-6\gamma_0}$$

(v) Given $V'(1) < 0$*, if we have* $b + 2g = 0$*, then* $\gamma_1 > 0$*, and the cubic equation* $V(x) = 0$ *has a repeated zero root, and a positive root greater than 1. Then, restricting p in the interval* $[0, 1]$*,* $V(p)$ *attains its maximum at*

$$p^* = \frac{2(a + 2f)}{2(a + 2f) + (a + 2f - 3)} < 1$$

Proof of Lemma A1 Suppose $V'(1) < 0$.

(i) Clearly $V'(1) < 0$ is equivalent to $a + 2f > 3$.

(ii) We now show that $\gamma_0 \geq 0$ would not be consistent with $V'(1) < 0$. If $\gamma_0 \geq 0$, then $2b + g \geq (a + 2f) - 1 > 2$ (because $a + 2f > 3$). But $2b + g > 2$ is not possible since $b < 1$ and $g < 0$. It follows that $\gamma_0 < 0$ whenever $V'(1) < 0$.

(iii) $\gamma_0 < 0$ implies $V(\infty) = -\infty$ and $V(-\infty) = \infty$. Condition $V'(0) = 2b + g > 0$ implies that $V(-\varepsilon) < 0$ for some small $\varepsilon > 0$.Then there must be a negative root and a positive root to the cubic equation $V(x) = 0$. Zero is the third root, because $V(0) = 0$. Thus the quadratic equation $V'(x) = 0$ must have two roots of opposite sign, and the positive one is smaller than 1, because $V'(1) < 0$.

(iv) $\gamma_0 < 0$ implies $V(\infty) = -\infty$ and $V(-\infty) = \infty$. Condition $V'(0) = 2b + g < 0$ implies that $V(\varepsilon) < 0$ for some very small $\varepsilon > 0$. Together with the fact that $V(1) > 0$, the cubic equation $V(x) = 0$ must have two positive real roots and a zero root. The quadratic equation $V'(x) = 0$ must have two positive real roots, their product being $(2b + g)/(-6\gamma_0) > 0$. Since $V'(1) < 0$, both of these real roots are inside the interval $(0, 1)$. The larger one yields a maximum for $V(p)$ when p is restricted to be in the unit interval $[0, 1]$.

(v) When $b + 2g = 0$, the function $V(x)$ becomes

$$V(x) = x^2 \left[a + 2f - x(a + 2f - 1)\right]$$

where $a+2f-1 > 0$ because of the hypothesis that $\gamma_0 < 0$. Then $V(\infty) = -\infty$ and $V(-\infty) = \infty$, and the cubic equation $V(x) = 0$ has the root $x = 0$ (repeated) and $x = (a+2f)/(a+2f-1) > 0$. The quadratic equation $V'(x) = 0$ has two roots, $x = 0$ and $x = \frac{2(a+2f)}{2(a+2f)+(a+2f-3)} < 1$.

Proposition 1 follows directly from Lemma A1.

Proof of Proposition 2

Now assume player ω becomes a Nashian. Then he has a dominant strategy, D, i.e. $p_\omega = 0$, i.e., $q_\omega = 1$. Then the expected payoff of player α is

$$\begin{aligned} V_\alpha(p_\alpha, p_{\beta,} 0) &= \left(f \times p_\alpha p_\beta q_\omega\right) + \left(g \times p_\alpha q_\beta q_\omega\right) + \left(b \times q_\alpha p_\beta q_\omega\right) \\ &= fp_\alpha p_\beta + gp_\alpha(1 - p_\beta) + b(1 - p_\alpha)p_\beta \end{aligned}$$

Similarly, the expected payoff of player β is

$$V_\beta(p_\alpha, p_{\beta,} 0) = fp_\alpha p_\beta + gp_\alpha(1 - p_\beta) + b(1 - p_\beta)p_\alpha$$

Clearly, since $b > f$ and $0 > g$, the payoff matrix facing players α and β has the property of a prisoner dilemma game if $f > 0$. If players α and β act according to Nash behavior, they will both choose D. However, if they collude and chooses a common mixed strategy, i.e. $p_\alpha = p_\beta = p$, the payoff of player i $(i = \alpha, \beta)$ will be

$$V_i(p) = (f - (b + g))p^2 + (b + g)p$$

Note that $V_i(0) = 0$ and $V_i(1) = f$.

The optimal p depends on whether $b + g < f$ or $b + g \geq f$. We consider three cases:

Case 1: $b + g < f$
Case 2: $b + g = f$
Case 3: $b + g > f$
Case 1: $V(p)$ is a strictly convex function. Therefore the maximum occurs at a corner. Since $V(0) = 0$ and $V(1) = 1$, the optimal choice is $p = 1$.
Case 2: $V(p) = p$. Clearly, the optimal choice is $p = 1$.

Case 3: Since $b+g>f$, $V(p)$ is a strictly concave function. Furthermore, $V'(0)=b+g>f>0$, and $V_i(\infty)<0$. Recalling that $V(0)=0$ and $V(1)=f>0$, we deduce that the function $V(p)$ must attain a maximum at some $\widetilde{p}>0$, where $\widetilde{p}$ defined by the FOC

$$(b+g)+2\,[f-(b+g)]\widetilde{p}=0$$

$$\widetilde{p}=\frac{(b+g)/2}{(b+g)-f}>0$$

Thus,

$$\arg\max_{p\in[0,1]}\begin{cases}\frac{(b+g)/2}{(b+g)-f} & \text{if } \frac{(b+g)/2}{(b+g)-f}<1\\ 1 & \text{if } \frac{(b+g)/2}{(b+g)-f}\geq 1\end{cases}$$

Thus the collusive equilibrium between players 1 and 2 arises iff

$$\frac{(b+g)}{2(b+g)-2f}<1$$

i.e.,

$$f<\frac{b+g}{2}$$

The intuition is that if the simple average of the off-diagonal payoff is greater than the payoff resulting from the strategy profile (C,C), then it pays to use a non-degenerate mixed strategy. At the equilibrium, the probability of playing C is

$$\frac{b+g}{(b+g)+(b+g-2f)}<1.$$

References

Arrow, K. J. (1973). Social responsibility and economic efficiency. *Public Policy, 21*, 303–317.

Bala, V., & Long, N. V. (2005). International trade and cultural diversity with preference selection. *European Journal of Political Economy, 21*(1), 143–162.

Bergstrom, T., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics, 29*(1), 25–49.

Brekke, K. A., Kverndokk, S., & Nyborg, K. (2003). An economic model of moral motivation. *Journal of Public Economics, 87*, 1967–1983.

Breton, M., Sbragia, L., & Zaccour, G. (2010). A dynamic model of international environmental agreements. *Environmental and Resource Economics, 45*, 25–48.

Buchholz, W. (2016). *Voluntary Public Good Provision as an Aggregative Game - A Pedagogical Note*. Typescript, University of Regensburg.
Buchholz, W., Cornes, R., & Rübbelke, D. (2014a). Potentially harmful international cooperation on global public good provision. *Economica, 81*(322), 205–223.
Buchholz, W., Falkinger, J., & Rübbelke, D. (2014b). Non-governmental public norm enforcement in large societies as a two-stage game of voluntary public good provision. *Journal of Public Economic Theory, 16*(6), 899–916.
Cornes, R., Long, N. V., & Shimomura, K. (2001). Drugs and pests: Intertemporal production externalities. *Japan and the World Economy, 13*(3), 255–278.
Cornes, R., & Sandler, T. (1984). The theory of public goods: Non-nash behavior. *Journal of Public Economics, 23*, 367–379.
Dockner, E., Jorgensen, S., Long, N. V., & Sorger, G. (2000). *Differential games, with applications to economics and management science*. Cambridge, U.K.: Cambridge University Press.
Grafton, Q., Kompas, T., & Long, N. V. (2016). *Virtual co-mover equilibria in games of climate change mitigation*. Paper presented at the 11th Tinbergen Institute-ERC Conference: Combating Climate Change, Amsterdam, April 2016.
Johansen, L. (1976). *The theory of public goods: Misplaced emphasis*. Oslo: Institute of Economics, University of Oslo.
Kay, J. (2015). *Other people's money. Masters of the universe or servants of the people*? London, U.K.: Profile Books Ltd.
Kemp, M. C. (1984). A note on the theory of international transfers. *Economics Letters, 14*(2–3), 259–262.
Kemp, M. C., & Long, N. V. (2009). Foreign aid in the presence of corruption: A differential game. *Review of International Economics, 17*(2), 230–243.
Laffont, J.-J. (1975). Macroeconomic constraints, economic efficiency and ethics: An introduction to Kantian economics. *Economica, 42*, 430–437.
Long, N. V. (2016), Kant-Nash equilibrium in a quantity-setting oligopoly. In P. von Mouche & F. Quartieri (Eds.), *Equilibrium theory for cournot oligopolies and related games: Essays in honour of Koji Okuguchi* (pp. 179–202). Switzerland: Springer.
Roemer, J. E. (2010). Kantian equilibrium. *Scandinavian Journal of Economics, 112*(1), 1–24.
Roemer, J. E. (2015). Kantian optimization: A microfoundation for cooperation. *Journal of Public Economics, 127*, 45–57.
Russell, B. (1945). *A history of western philosophy*. London: George Allen and Unwin.
Seabright, P. (2010). *The company of strangers: A natural history of economic life*. Princeton, NJ: Princeton University Press.
Sen, A. (1977). Rational fools: A critique of the behavioral foundation of utility. *Philosophy and Public Affairs, 6*(4), 317–344.
Sen, A. (1993). Does business ethics make economic sense? *Business Ethics Quarterly, 3*(1), 45–54.
Sethi, R., & Somanathan, E. (1996). The evolution of social norms in common property resource use. *American Economic Review, 86*(4), 766–788.
Smith, A. (1790 [2002]) *The theory of moral sentiments* (revised edition, 1790). Republished in the series Cambridge Texts in the History of Philosophy, edited by Knud Haakonssen.
Warr, P. (1983). The private provision of public goods is independent of the distribution of income. *Economics Letters, 13*(2–3), 207–211.

Social Creation of Pro-social Preferences for Collective Action

Avinash Dixit and Simon Levin

1 Introduction and Motivation

Study of collective action to provide public goods was the focus of much of Richard Cornes' work. Attainment of aggregate efficiency in these situations has to overcome free riding by selfish participants. Most of the work in this area, including the classic book of Cornes and Sandler (1996), was grounded in economists' traditional assumption of exogenous and self-regarding preferences. Cornes's occasional excursions into other-regarding preferences involved goods with joint private and public characteristics (e.g. Cornes and Sandler 1996, Chap. 8), and intra-family altruism for transfers (e.g. Cornes and Silva 1999) or for public good provision (e.g. Cornes et al. 2012). Economics in recent years has increasingly recognized that people have pro-social preferences in larger social groups, and is beginning to recognize that preferences are not exogenous but are socially formed. In this paper we develop a model with these features, and examine to what extent such pro-socialness can be instilled and help solve collective action problems.

Pro-social preferences and other-regarding behaviors more generally are a fact of life, though it is often puzzling how they are sustained (Henrich et al. 2001; Gintis 2003; Fehr and Gintis 2007; Akcay et al. 2009; Henrich et al. 2010). The most plausible explanation will combine genetic and evolutionary pathways with socio-cultural processes to incentivize and reinforce pro-sociality. In this paper we focus on one such societal process. Our basic framework builds on earlier work by the first author (Dixit 2009). The framework is a general one, where individuals allocate their efforts or resources between their own interests and the public good. The analysis applies equally to investments that limit the damage to common pool

A. Dixit (✉) • S. Levin
Department of Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA
e-mail: dixitak@princeton.edu

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_7

resources; examples are fisheries management, climate change and effectiveness of antibiotics (Hardin 1968; Ostrom 1990; Levin 1999; Smith et al. 2005).

In evolutionary biology, individuals are endowed with genes that determine the phenotypes that in turn determine their strategies, and fitter strategies proliferate faster. In social settings, individuals are indeed born with some unchangeable behaviors, but they acquire many other behaviors during a long period of socialization that begins with families, extends for many years in schools, and continues at various levels of intensity into adulthood and indeed throughout life. The early years of life, when children are most impressionable and their preferences and behavior can be molded substantially, should be the most crucial phase in this long process.

Developmental psychologists have studied the process in more detail. Hoffman (2000, pp. 10–11) describes it thus: "When a child experiences, repeatedly, the sequence of transgression followed by a parent's induction[1] followed by child's empathetic distress and guilt feeling, the child forms Transgression → Induction → Guilt scripts . . . [That] may become strong enough with repetition, and when combined with cognitive development and peer pressure . . . may be effective. That is, peer pressure *compels* children to realize that others have claims; cognition *enables* them to understand others' perspectives; empathic distress and guilt *motivate* them to take others' claims and perspectives into account." (Emphasis in the original.) The importance of peer pressure, and an important role for schools (or similar settings where many unrelated children interact), is emphasized in sociological literature. Thus Boocock and Scott (2005, p. 84) find that "on a wide range of attitudes and behaviors, kids tend to become more like their friends and less like their parents."

Our purpose in this paper is to construct a simple model of such collective action to instill pro-social preferences in children.[2] We emphasize again that we do not claim this as the only or even the predominant way in which societies generate or sustain pro-sociality; it can coexist with other pathways of genetic or social evolution. But it is clearly a significant one in reality, and worth analysis and theoretical exploration on its own.

The model focuses on public good provision. Final output is produced using the public good and private effort. The public good increases the productivity of private effort. Therefore when each person contributes to the public good, this raises all individuals' personal benefit, which we call their private or *selfish utility.* When deciding whether and how much to contribute to the public good, purely selfish individuals would ignore the benefits that flow to others. Therefore they would contribute too little, to the detriment of all. This is the standard prisoners' dilemma

[1]Induction in this context is a mild form of discipline technique. Hoffman defines and explains it as follows: "When children harm or are about to harm someone—the parent, a sibling, a friend— . . . indicate[s] implicitly or explicitly that the act is wrong and that the child has committed an infraction. . . . This creates the condition for feeling empathy-based guilt. Hoffman (2000, pp. 150–151).

[2]The need for collective action for preference formation is the crucial respect in which our model differs from other models where individual parents shape their children's preferences, for example Bisin and Verdier (2001), Tabellini (2008).

of collective action.[3] One way to ameliorate it is for individuals to have pro-social preferences; we call such an objective the individual's *pro-social utility*.

Each parent cares about her children's welfare. Parents with such concerns would like the next generation to solve its prisoners' dilemma of public good provision; therefore they would like the next generation to have pro-social preferences. Of course any one parent would not want its child to become the only pro-social actor if all other children grow up to be selfish; that would make the child a sucker in a prisoners' dilemma game. However, each parent may be willing to vote for a tax that finances education for all children to instill pro-social preferences in all.[4] This is how we model the process. It fits with what we observe in reality: schools do devote a substantial amount of time and resources to socialization and to teaching concepts of civic duties, concern for and responsibility toward others, social norms of behavior, and so on.

In other words, what parents are giving their children is not a contribution to the public good directly as in Cornes et al. (2012), but altruistic preferences that will enable the children's generation to solve its own collective action problem more efficiently.

The model is an extended numerical example to develop the ideas. We use special functional forms and parameter values chosen to facilitate solution. But the intuitions it creates are appealing and the qualitative results should remain valid in more general conditions.

2 Equilibrium with Given Pro-socialness

Begin with one generation, and examine how pro-social preferences improve the provision of the public good and the utilities of individuals. There are n individuals, labeled $i = 1, 2, \ldots n$. Each can exert two types of effort: private x_i, and public z_i. The public good may consist of the effort itself, for example volunteered time, or it may be a good or service produced one-for-one using aggregate public effort; either interpretation works equally well. Denote the average public effort by

$$\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i \,. \tag{1}$$

[3]The problem of the commons, where the issue is how to dissuade individuals from creating a public bad, is the mirror image of this, and can be analyzed by similar methods. In other contexts, sufficiently pro-social preferences may lead to a coordination or "assurance" game. Such more complex interaction topologies are subject of our ongoing work.

[4]See Friedman (1962, p. 191) for an early argument along these lines in favor of government action to alleviate poverty.

Then the income of individual i is given by

$$y_i = (1 + \bar{z})\, x_i\ . \tag{2}$$

Thus a higher $\bar{z}$ raises the (average and) marginal product of each individual's private effort, as mentioned above. By using the average $\bar{z}$, not the total $\sum_{i=1}^{n} z_i$, in (1), we are assuming that the public effort is a public good with congestion. The no-congestion case is worth separate analysis.

The private or selfish utility of i is assumed to be

$$u_i = y_i - \tfrac{1}{3}\,(x_i + z_i)^2\ . \tag{3}$$

If individuals are selfish, the Nash equilibrium of their non-cooperative choices has no public effort:

$$x_i = 3/2, \qquad z_i = 0, \qquad y_i = 3/2, \qquad u_i = 3/4\ . \tag{4}$$

The calculation, and similar calculations to follow, are simple but sometimes tedious, and do not contribute to the intuition; therefore they are relegated to the Appendix, section "Nash Equilibrium with Pro-sociality".

Contrast this with the symmetric cooperative optimum. Here we choose common effort levels $x_i = x$ and $z_i = z$ for all i. Then $\bar{z} = z$, and the common utility level $u_i = u$ for all i is given by

$$u = x\,(1 + z) - \tfrac{1}{3}\,(x + z)^2 \tag{5}$$

We show in the Appendix, section "Nash Equilibrium with Pro-sociality" that the optimum that maximizes u is given by

$$x_i = 2, \qquad z_i = 1, \qquad y_i = 4, \qquad u_i = 1\ . \tag{6}$$

The contribution to public effort raises the incentive to make private effort, leading to much higher incomes, sufficiently higher to overcome the disutility of the greater effort.

Some societies may achieve this optimum by command and control. But this is often infeasible in open democratic societies; and even if feasible, many would prefer gentler methods. The one we consider is to change preferences to include a pro-social element. Begin by examining the effect of pro-social preferences and then consider how they can be instilled. Suppose individual 1's pro-social utility is

$$v_1 = u_1 + \gamma \sum_{i=2}^{n} u_i\ , \tag{7}$$

where the parameter γ is assumed to be in the interval $(0, 1)$, and captures the idea that each individual includes some concern for the welfare of other individuals, but not as much as for her own. The pro-social utility of individuals $2, 3, \ldots n$ is defined similarly. If

$$\gamma \leq \frac{2n-3}{3(n-1)}\,, \tag{8}$$

the Nash equilibrium is the same as in (4), with $z_i = 0$ for all i. The right-hand side of (8) is the minimum threshold of pro-social preference needed to induce positive public effort. Thus just a little pro-socialness does not work; this is similar to the result of Rabin (1993) in the context of fairness. The threshold rises with n, but goes to 2/3, not 1, as $n \to \infty$: even in very large societies, the threshold is consistent with regarding others' utility worth less than one's own.

Use the abbreviation

$$\phi = \frac{1 + \gamma\,(n-1)}{n}\,. \tag{9}$$

With the assumption $\gamma < 1$, we have $\phi < 1$. For large n, $\phi = \gamma$ approximately. Then (8) becomes simply $\phi \leq 2/3$. Most public good situations involve large populations. Therefore for most of what follows we express all formulas in terms of ϕ and think of it as the pro-socialness parameter.

If $\phi > 2/3$, the symmetric Nash equilibrium of individual contributions to public effort is:

$$x_i = \frac{2}{2-\phi}, \quad z_i = \frac{3\phi - 2}{2-\phi}, \quad y_i = \frac{4\phi}{(2-\phi)^2}, \quad u_i = \frac{\phi\,(4-3\phi)}{(2-\phi)^2}\,. \tag{10}$$

The resulting pro-social utilities are

$$\begin{aligned} v_i &= [1 + (n-1)\,\gamma]\, u_i = n\,\phi\, u_i \\ &= n\,\frac{\phi^2\,(4-3\phi)}{(2-\phi)^2}\,. \end{aligned} \tag{11}$$

As ϕ increases from 2/3 to 1, (10) moves monotonically from the purely selfish (4) to the optimal (6). Therefore in this range, if everyone has more pro-social preferences, that raises everyone's *selfish* utility. Figure 1 shows this functional relationship.

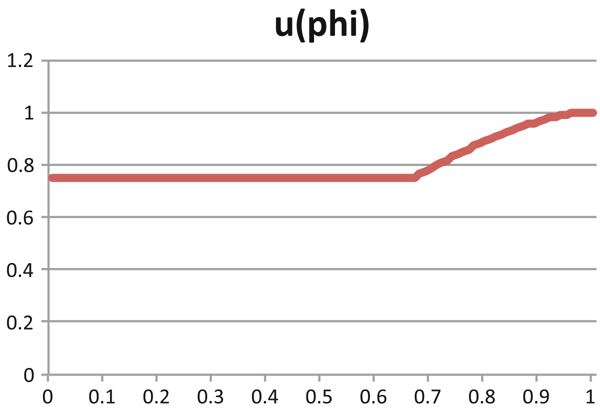


Fig. 1 Selfish utility u in Nash equilibrium as function of pro-socialness parameter ϕ

3 Choosing Children's Pro-socialness

Now introduce a succession of generations. Each individual has one child. Each parent cares about the utility of her child. All parents in a cohort recognize that if their children's generation had enough pro-sociality, they would be able to achieve higher utility as in the previous section. The parents would be willing to sacrifice some of their own utility to achieve such an outcome. But that requires coordinated action—just one parent bearing a cost to educate her own child to be pro-social would merely convey the benefit of her child's public effort to the children of other parents. We model the coordinated action as choosing a tax levied on all members of the parent generation to fund education that instills pro-sociality in the children's generation. We assume all individuals in a generation to be identical, therefore their choice is unanimous and made to maximize the utility (which includes the concern for the child's well-being) for each of them.

We consider two cases. In the first, this process happens just once. Generation 1, which had no pro-sociality instilled into it by generation 0, wakes up and realizes that it can improve the well-being of generation 2 through the socializing education. Generation 1 does not know whether generation 2 will continue to pursue any such actions, and ignores that possibility. This one-off model may be a purely hypothetical thought-experiment, but it serves a useful purpose of introducing the basic idea.

The other case is a steady state. At some time before the action starts, a cohort or generation of "founding mothers" meets to write a constitution. This specifies the tax to be levied on each member of every generation from that point on to educate the next generation. We leave aside the problem of what happens to the very first generation, or assume that this generation of the founding mothers somehow chooses to build the right level of pro-socialness into itself in course of their constitutional deliberations. Thus we assume that the steady state is attained at once and continues; we do not examine dynamics starting from an arbitrary initial condition and the possibility that it may not converge to a steady state.

3.1 One-Off Action

Write u^a for the selfish utility of any one adult in the generation that decides to educate its children for pro-sociality, and u^c for that of her child; each is defined as in (3). The adult's overall utility including that of her child's well-being is defined by

$$U^a = u^a + \delta\, u^c\ , \tag{12}$$

which includes concern for one's own child but not for anyone else. We assume $0 < \delta < 1$.

Education can give the child a social utility with parameter ϕ, related to the γ as in (9). We assume the following form for the cost of this per capita:

$$t = \frac{k}{1-\phi} \qquad \text{for } \phi > 0, \quad \text{or} \quad \phi = 1 - \frac{k}{t} \qquad \text{for } t > k. \tag{13}$$

Thus k is a fixed cost for imparting any positive level of pro-socialness, however small. As ϕ increases from o to 1, t is an increasing convex function of ϕ, and $t \to \infty$ as $\phi \to 1$: the marginal cost of preference-formation is increasing, and it is infinitely costly to make each individual fully internalize social welfare.

Endowed with this ϕ, the children's generation will achieve a Nash equilibrium of their game of effort allocation between private and public uses, and that will yield its u^c as a function $u^c(\phi)$, as defined in Eqs. (4) and (10).

The parent generation had no pro-socialness installed in it; therefore its u^a is determined by the Nash equilibrium of the game with $\gamma = 0$, namely $u^a = 0.75$ as we found in Sect. 2. The tax is subtracted from this. Also, in our thought-experiment of this subsection, the parent generation acts as if the children's generation will not organize any pro-socializing effort for the generation to follow, i.e. for this generation's grandchildren. Accordingly, no tax is anticipated or subtracted from u^c. Therefore this generation chooses t, or equivalently ϕ, to maximize U^a. Using (4) and (10), this can be expressed as a function of ϕ:

$$U^a \equiv f(\phi) = \begin{cases} \dfrac{3}{4} - \dfrac{k}{1-\phi} + \delta\, \dfrac{3}{4} & \text{for } \phi < 2/3 \\[2ex] \dfrac{3}{4} - \dfrac{k}{1-\phi} + \delta\, \dfrac{\phi\,(4-3\phi)}{(2-\phi)^2} & \text{for } 2/3 < \phi < 1 \end{cases} \tag{14}$$

The details of the solution are in the Appendix, section "One-Time Education for Pro-socialness". To state it more compactly, define $\theta = (k/\delta)^{1/3}$. Then there is a critical level $\theta^* \approx 0.305$ such that when $\theta > \theta^*$ (corresponding to $k > 0.028\,\delta$ approximately), it is optimal to choose $\phi = 0$. If $\theta < \theta^*$, that is, $k < 0.028\,\delta$, the optimum choice is

$$\phi = \frac{2\,(1-\theta)}{2-\theta}\,. \tag{15}$$

Thus, if the cost of education is low relative to the regard for children's welfare, the parent generation will instill pro-social preferences in the next generation through education. The level of ϕ that is instilled when k is just below its threshold which we saw above to be approximately $0.028\,\delta$ (that is, when θ is just below it threshold θ^* approximately 0.305) is approximately 0.82, which significantly exceeds the threshold $\phi = \frac{2}{3}$ needed to induce a small positive public effort. This jump is due to the fixed cost feature of the education technology. If the fixed cost k becomes negligibly small, θ goes to zero and ϕ goes to 1; in this limit the full social optimum can be approached.

Since $\delta < 1$, $k \geq 0.028$ is sufficient to rule out pro-socialness in the one-off situation. This will be contrasted with the steady-state case that follows.

3.2 Steady State

Now we consider the steady state of an ongoing succession of generations, where a stationary policy of taxation and education is chosen in advance in a binding constitution. The procedure is similar to that in Arrow and Levin (2009). Let $v(n)$ denote the utility of any one member of generation n, including the effects of the Nash equilibrium it achieves given its own pro-sociality and the tax it pays to educate generation $n+1$, but not the utility this person gets from the well-being of its own child. Let $V(n)$ denote the comprehensive utility of this person including this parental concern also. Then we have

$$V(n) = v(n) + \delta\, V(n+1) \tag{16}$$

We assume a steady state in which the constitution fixed in advance also fixes the tax to be the same for all generations, and then the equilibrium and utilities are the same for all generations n. Denoting these common values by v and V, we have

$$V = \frac{v}{1-\delta} \tag{17}$$

Maximizing V is therefore equivalent to maximizing v, independently of δ. The extent of each generation's concern for its children does not matter in the steady state; the founding mothers have taken it into account and made the best common choice for all.

If u is any one individual's utility and t the tax, we have

$$v = [1 + (n-1)\gamma]\,(u - t) = n\,\phi\,(u - t) \tag{18}$$

Observe that the pro-socialness instilled into a parent leads her to include the effects both of the utility and the tax on her contemporaries.

Using Eqs. (4), (11), and (13) gives the objective:

$$v = \begin{cases} n\,\phi\,\left[\dfrac{3}{4} - \dfrac{k}{1-\phi}\right] & \text{for } \phi < 2/3 \\ n\,\phi\,\left[\dfrac{\phi(4-3\phi)}{(2-\phi)^2} - \dfrac{k}{1-\phi}\right] & \text{for } 2/3 < \phi < 1 \end{cases} \tag{19}$$

A combination of analytical and numerical methods detailed in the Appendix, section "Education for Pro-socialness in Steady State" yields the following solution for the choice of ϕ to maximize v:

If $k < 1/12$ $(= 0.0833)$, the function is increasing throughout the range $0 < \phi < 2/3$, and then has a peak in the range $2/3 < \phi < 1$. Thus it has a unique optimum in the range where sufficient pro-sociality is instilled to elicit public effort. Figure 2

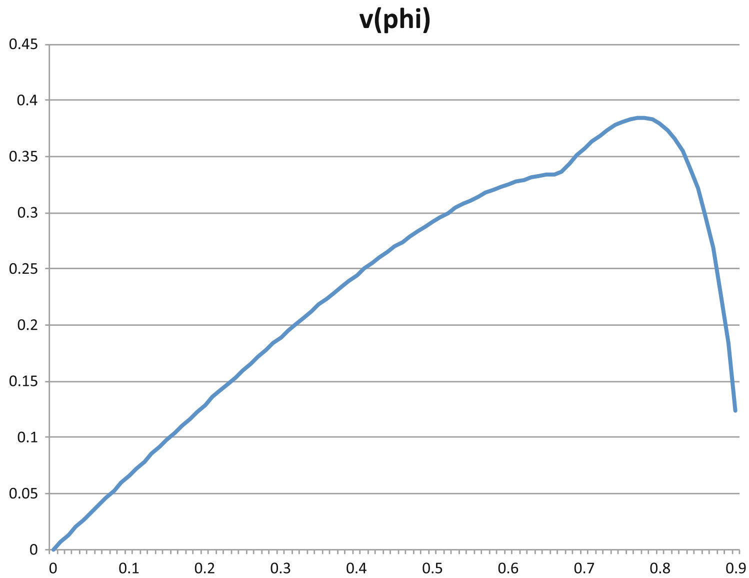
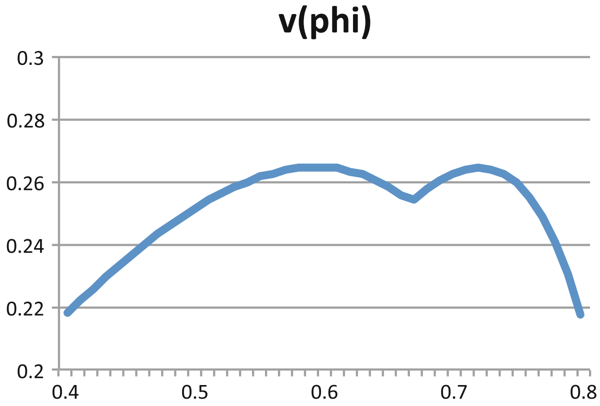


Fig. 2 Steady state: case of low k

Fig. 3 Steady state: case of mid-range k



graphs this for the upper extreme value of the range, $k = 1/12$, where the function just becomes flat approaching $\phi = 2/3$ from the left.

If $(0.0833 =)\ 1/12 < k < 1/6\ (= 0.1667)$, the function has local peaks in each of the ranges $0 < \phi < 2/3$ and $2/3 < \phi < 1$. The former yields the optimum if $k < 0.1235$ and the latter if $k > 0.1235$. Figure 3 graphs this in the borderline case of $k = 0.1235$.

If $(0.1667 =)\ 1/6 < k < 3/4$, the function has a peak in the range $0 < \phi < 2/3$ and is decreasing throughout the range $2/3 < \phi < 1$. Therefore it has a unique optimum in the former range, where some pro-sociality is instilled but not enough to elicit any public effort. We comment on this below. Figure 4 graphs the function for the lower extreme value of the range, $k = 1/6$, when the function is flat just to the right of $\phi = 2/3$.

For $k > 3/4$, the function is decreasing throughout, and $\phi = 0$ is the optimum.

Some features of the solution require comment: (1) pro-sociality is instilled for a much larger range of k in the steady state than in the one-off case. Recall that in the latter, $k = 0.028\,\delta < 0.028$ was the upper limit of the cost parameter for which enough pro-sociality to induce public effort would be instilled. Now that limit is $k = 0.123$. (2) As remarked earlier, the extent of pro-sociality is independent of δ. The founding mothers once and for all figure out what is best for every generation and specify it as a rule in the constitution. (3) If $0.123 < k < 0.75$, some pro-sociality is instilled but not enough to lead to any public effort. This is because the inclusion of other people's utility in each person's own social utility has value in itself. For example, if $k = 0.125$, it turns out that the optimum is $\phi = 0.6$. Then $t = 0.125/(1 - 0.6) = 0.31$ and each person's selfish utility is $u = 0.75 - 0.31 = 0.44$. But the social utility is $v = 0.44 + (n - 1)\,\gamma \times 0.44 = n\,\phi \times 0.44$. (4) If k increases across 0.1235, the optimum undergoes a discrete downward jump, in Fig. 3 from about $\phi = 0.72$ to 0.59, and in the Nash equilibrium the extent of public effort drops discretely to zero. As in the one-off case, the reason is the fixed cost nature of the education technology.

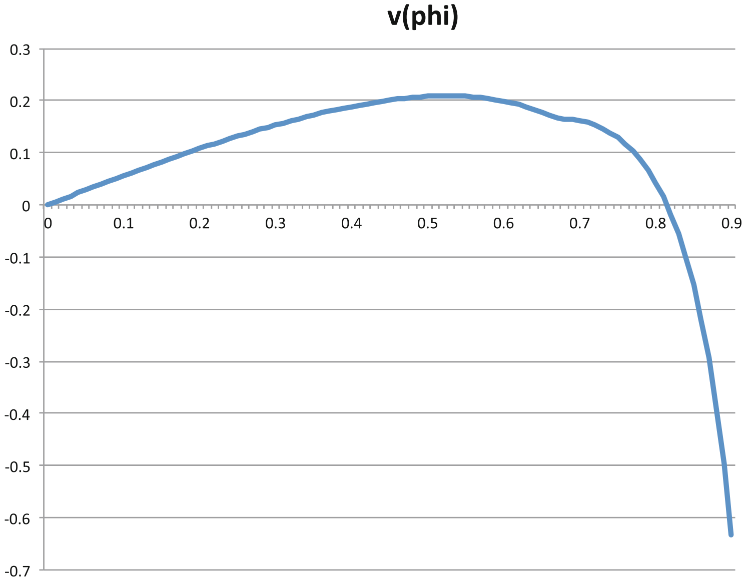


Fig. 4 Steady state: case of high k

4 Concluding Comments

Most societies put considerable effort into socializing youngsters to achieve better behavior and outcomes. In this paper we developed a simple mathematical model of this process, and examined its implications for voluntary provision of public goods. We hope this will prove a useful step in the endeavor of incorporating insights not only from psychology but also from sociology or social psychology into economics, and improve our understanding of how collective action is organized. We hope thereby to help bridge the focus of Richard Cornes's lifetime interest with recent advances in behavioral and sociological economics.

Acknowledgements Funding was provided by the National Science Foundation grants GEO-1211972 and OCE-1426746, and by the Nordforsk-funded project Green Growth Based on Marine Resources: Ecological and Socio-Economic Constraints (GreenMAR) to Simon A. Levin.

Appendix: Mathematical Derivations

Nash Equilibrium with Pro-sociality

Begin with the purely selfish case, set out in Sect. 2, Eqs. (1)–(3). The Kuhn-Tucker conditions for individual i's choice of (x_i, z_i) to maximize u_i are

$$\frac{\partial u_i}{\partial x_i} = 1 + \overline{z} - \tfrac{2}{3}\,(x_i + z_i) \leq 0, \quad x_i \geq 0\,, \tag{20}$$

$$\frac{\partial u_i}{\partial z_i} = \tfrac{1}{n}\,x_i - \tfrac{2}{3}\,(x_i + z_i) \leq 0, \quad z_i \geq 0\,, \tag{21}$$

with complementary slackness in each equation. Note that in (21) we have used $\partial\overline{z}/\partial z_i = 1/n$. The matrix of second-order partials is

$$\begin{pmatrix} -\frac{2}{3} & -\frac{2}{3} + \frac{1}{n} \\ -\frac{2}{3} + \frac{1}{n} & -\frac{2}{3} \end{pmatrix}.$$

The diagonal elements of this are negative, and the determinant is

$$\left(\tfrac{2}{3}\right)^2 - \left(\tfrac{2}{3} - \tfrac{1}{n}\right)^2 > 0\,.$$

Therefore the matrix is negative definite, so the second-order sufficient conditions are met and the Kuhn-Tucker conditions yield the global maximum of u_i. (This will continue to be so in all the variants of the model considered here, and will not be mentioned further.)

First try the purely selfish solution where $x_i > 0$ and $z_i = 0$ for all i. Then $\overline{z} = 0$ also, and the conditions (20) and (21) are

$$1 - \tfrac{2}{3}\,x_i = 0, \qquad \left(\tfrac{1}{n} - \tfrac{2}{3}\right)\,x_i \leq 0\,,$$

or

$$x_i = \tfrac{3}{2}, \qquad \tfrac{1}{n} \leq \tfrac{2}{3}\,.$$

The inequality is true for $n \geq 2$. Therefore the solution $x_i = 3/2$, $z_i = 0$ in (4) is verified. The y_i and u_i are easily computed.

Next consider the symmetric social optimum. Recall that the common private and public effort levels $x_i = x$ and $z_i = z$ for all i, are chosen to maximize the common utility level

$$u = x\,(1 + z) - \tfrac{1}{3}\,(x + z)^2\,.$$

The first-order conditions are

$$\frac{\partial u}{\partial x} = 1 + z - \tfrac{2}{3}\,(x+z) = 0\,,$$
$$\frac{\partial u}{\partial z} = x - \tfrac{2}{3}\,(x+z) = 0\,.$$

These yield the solution $x = 2$, $z = 1$ in (6). The resulting y and u are easily found.

Next consider equilibria where people have the pro-social utility (7), with the same γ for all. The Kuhn-Tucker conditions for person 1 are:

$$\frac{\partial v_1}{\partial x_1} = 1 + \bar{z} - \tfrac{2}{3}\,(x_1 + z_1) \le 0, \quad x_1 \ge 0\,, \tag{22}$$

$$\frac{\partial v_1}{\partial z_1} = \tfrac{1}{n}\,x_1 - \tfrac{2}{3}\,(x_1 + z_1) + \gamma \sum_{j=2}^{n} \tfrac{1}{n}\,x_j \le 0, \quad z_i \ge 0\,, \tag{23}$$

with complementary slackness in each. Similar conditions obtain for the other individuals.

See if the selfish solution with $x_i > 0$, $z_i = 0$ still works. The conditions (22) and (23) become

$$1 - \tfrac{2}{3}\,x_i = 0, \qquad \tfrac{1}{n}\,x_1 - \tfrac{2}{3}\,x_1 + \gamma \sum_{j=2}^{n} \tfrac{1}{n}\,x_j \le 0\,,$$

or

$$x_i = \tfrac{3}{2}, \qquad \tfrac{3}{2}\left(\tfrac{1}{n} - \tfrac{2}{3} + \gamma\,\tfrac{n-1}{n}\right) \le 0\,.$$

The inequality becomes

$$\phi = \frac{1 + (n-1)\,\gamma}{n} \le \frac{2}{3}\,,$$

which is equivalent to (8) in the text.

When this condition is not met, look for a symmetric Nash equilibrium with $x_i = x > 0$ and $z_i = z > 0$ for all i. The conditions (22) and (23) become

$$1 + z = \tfrac{2}{3}\,(x+z)\,,$$
$$\phi\,x = \tfrac{2}{3}\,(x+z)\,.$$

These yield the solution (10) in the text.

Writing u for the common level of selfish utility, it is then mechanical to verify

$$\frac{du}{d\phi} = \frac{8\,(1-\phi)}{(2-\phi)^2}\,.$$

Therefore u is an increasing function of ϕ over the range $(\frac{2}{3}, 1)$. Thus more pro-socialness achieves higher selfish utilities all round.

One-Time Education for Pro-socialness

In the range $0 \le \phi \le 2/3$, it is obviously best to set $\phi = 0$ and get $U^a = 0.75\,(1+\delta)$. In the range $2/3 \le \phi \le 1$, use $\theta = (k/\delta)^{1/3}$, to substitute $k = \delta\,\theta^3$, and write the formula defining the function for $\phi \ge \frac{2}{3}$ as

$$f(\phi) = 0.75 + \delta\left[\frac{\phi\,(4-3\phi)}{(2-\phi)^2} - \frac{\theta^3}{1-\phi}\right]. \tag{24}$$

It is then mechanical to verify

$$f'(\phi) = \frac{\delta}{(1-\phi)^2}\left[8\left(\frac{1-\phi}{2-\phi}\right)^3 - \theta^3\right].$$

Therefore

$$f'(\phi) > 0 \quad \text{iff} \quad 2\,\frac{1-\phi}{2-\phi} > \theta, \quad \text{i.e.} \quad \phi < \phi^* = \frac{2\,(1-\theta)}{2-\theta}\,. \tag{25}$$

Therefore $f(\phi)$ is single-peaked, and its maximum occurs where $f'(\phi) = 0$, that is, at $\phi = \phi^*$. Substituting and simplifying, the maximum value is

$$f(\phi^*) = 0.75 + \delta\left(\theta^3 - 3\,\theta^2 + 1\right).$$

If this exceeds $0.75\,(1+\delta)$, then ϕ^* maximizes $f(\phi)$ in (14); otherwise $\phi = 0$ yields the maximum of $f(\phi)$.

We also need to restrict $\phi^* > 2/3$ to have an equilibrium that results in the utilities that enter the construction of $f(\phi)$. From the definition in (7), we see that $1 \ge \phi^* > 2/3$ corresponds to $0 \le \theta < 1/2$. Now define

$$h(\theta) = \theta^3 - 3\,\theta^2 + 1 - 0.75 = \theta^3 - 3\,\theta^2 + 0.25\,.$$

We have

$$h'(\theta) = 3\,\theta^2 - 6\,\theta = 3\,\theta\,(\theta - 2)\,,$$

which is negative over the interval $(0, \frac{1}{2})$. Therefore $h(\theta)$ is a decreasing function throughout this range. Also $h(0) = 1/4 > 0$ and $h(\frac{1}{2}) = -3/8 < 0$. Therefore there is a unique θ^* in the interval such that $h(\theta) > 0$ for $\theta < \theta^*$ and $h(\theta) < 0$ for $\theta > \theta^*$. Numerical calculation shows that $\theta^* \approx 0.305$. This completes the proof of the statements in the text leading to (15).

Education for Pro-socialness in Steady State

Write the objective function in (26) as $v = n\,g(\phi)$ where

$$g(\phi) = \begin{cases} 0.75\,\phi - \frac{k\,\phi}{1-\phi} & \text{for } \phi < 2/3 \\ \frac{\phi^2\,(4-3\phi)}{(2-\phi)^2} - \frac{k\,\phi}{1-\phi} & \text{for } 2/3 < \phi < 1 \end{cases} \tag{26}$$

For $0 \le \phi < 2/3$ we have

$$g'(\phi) = 0.75 - \frac{k}{(1-\phi)^2}$$

and

$$g''(\phi) = \frac{-2\,k}{(1-\phi)^3} < 0$$

Therefore $g(\phi)$ is concave in this range. Also

$$g'(0) = 0.75 - k > 0 \quad \text{if } k < 0.75$$

and

$$g'(2/3) = 0.75 - 9\,k < 0 \quad \text{if } k > 1/12$$

For $2/3 < \phi < 1$ we have, after some tedious algebra,

$$g'(\phi) = \frac{\phi\,(16 - 18\,\phi + 3\,\phi^2)}{(2-\phi)^3} - \frac{k}{(1-\phi)^2}$$

Then

$$g'(2/3) = 3/2 - 9\,k > 0 \quad \text{if } k < 1/6$$

and

$$g'(\phi) \to -\infty \quad \text{as} \quad \phi \to 1$$

In general, $g(\phi)$ is not concave throughout the range $2/3 < \phi < 1$. (Specifically, the first term on the right hand side of (26) is convex for $2/3 < \phi < 4/5$ and concave for $4/5 < \phi < 1$, so $g(\phi)$ can be convex for a sub-range to the right of 2/3 if k is small.) But numerical calculation shows that

$$(1-\phi)^2\, g'(\phi) = \frac{\phi\,(1-\phi)^2\,(16 - 18\,\phi + 3\,\phi^2)}{(2-\phi)^3} - k$$

is a decreasing function of ϕ. It equals $1/6 - k$ for $\phi = 2/3$ and $-k$ for $\phi = 1$. Therefore, if $k < 1/6$, $g'(\phi)$ is positive for $\phi <$ a critical value ϕ^* (which is uniquely defined for each k and of course depends on k), and negative for $\phi > \phi^*$. Then $g(\phi)$ is increasing for $2/3 < \phi < \phi^*$ and decreasing for $\phi^* < \phi < 1$, i.e. $g(\phi)$ has a unique interior maximum at ϕ^* in the range $2/3 < \phi < 1$.

Putting all this information together gives the statements in the text.

References

Akcay, E., Van Cleve, J., Feldman, M. W., & Roughgarden, J. (2009). A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 19061–19066.

Arrow, K. J., & Levin, S. A. (2009). Intergenerational resource transfers with random offspring numbers. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 13702–13706.

Bisin, A., & Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory, 97*(2), 298–319.

Boocock, S. S., & Scott, K. A. (2005). *Kids in context: The sociological study of children and childhoods*. Lanham, MD: Rowan and Littlefield.

Cornes, R., Itaya, J.-I., & Tanaka, A. (2012). Private provision of public goods between families. *Journal of Population Economics, 25*(4), 1451–1480.

Cornes, R., & Sandler, T. (1996). *The theory of externalities, public goods, and club goods* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cornes, R., & Silva, E. C. D. (1999). Rotten kids, purity, and perfection. *Journal of Political Economy, 107*(5), 1034–1040

Dixit, A. (2009). Governance institutions and economic activity. *American Economic Review, 99*(1), 5–24.

Fehr, E., & Gintis, H. (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology, 33*, 43–64.

Friedman, M. (1962). *Capitalism and freedom*. Chicago, IL: University of Chicago Press.

Gintis, H. (2003). Solving the puzzle of pro-sociality. *Rationality and Society, 15*, 155–187.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2001). In search of Homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review, 91*, 73–78.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–135.

Hoffman, M. L. (2000). *Empathy and moral development*. Cambridge, UK: Cambridge University Press.

Levin, S. A. (1999). *Fragile dominion: Complexity and the commons*. Reading, MA: Perseus Books.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.

Rabin, M. (1993). Incorporating fairness into game-theory and economics. *American Economic Review, 83*, 1281–1302.

Smith, D. L., Levin, S. A., & Laxminarayan, R. (2005). Strategic interactions in multi-institutional epidemics of antibiotic resistance. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 3153–3158.

Tabellini, G. (2008). The scope of cooperation: Values and incentives. *Quarterly Journal of Economics, 123*(3), 905–950.

Community Size and Public Goods in Hume's *Treatise of Human Nature*

Simon Vicary

David Hume was a central figure of the Scottish Enlightenment of the eighteenth century. It has long been recognised that his thoughts on economics and government were influential in the establishment of economics, or rather political economy, as a distinctive area of enquiry, particularly through his influence on Adam Smith. However, in recent years his writings have attracted attention in their own right, and have been read for more than historical interest. Indeed, he has gained something of a reputation as the economist's philosopher. Many different themes have been explored. Some are philosophical. For example, Sturn (2004) analyses Hume's notion of utility in relation to that used in modern economics, whilst Sugden (2005, 2006) and Dow (2009), pursuing parallel themes, emphasise the question of rationality, and how his ideas relate to modern debates on the meaning of rationality and individual motivation. The latter writers make a clear link between Hume's ideas and modern behavioural economics. Leaving aside his essays on money and the balance of payments, his work has also been linked to a rather different strand of economics. According to Binmore (2005) he was "the original inventor of reciprocal altruism- the first person to recognise that the equilibrium ideas now studied in game theory are vital to an understanding of how human societies work". In similar vein, Vanderschraaf (1998) explores in detail the close relationship between Hume's thinking on justice and modern developments in formal game theory. Hardin (2007) goes further in placing Hume's game theory at the very centre of his thinking on moral and political questions. He claims that many criticisms of Hume made before

I am grateful to a referee as well as to Ted Bergstrom, Marisa Ratto and Mich Tvede for helpful comments on an earlier version of this paper.

S. Vicary (✉)
Department of Economics, Newcastle University Business School, University of Newcastle-upon-Tyne, 5, Barrack Road, Newcastle upon Tyne, NE1 4SE, UK
e-mail: simon.vicary@newcastle.ac.uk

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_8

the wider understanding of game theory that came in the late twentieth century simply reflect a misunderstanding of this aspect of his work.

The ideas for which Hume is now celebrated in economics often relate in some way to some form of collective benefit, or what might in broad terms be thought of as a public good. Both Levy (1984) and Nuyen (1986) argue convincingly that Hume's concept of justice, a central concern of Volume Three of his *Treatise on Human Nature*, is best thought of as a public good. It has long been understood that public goods in some sense played a part in Hume's thinking. His example of two farmers wishing to drain a meadow has been widely quoted as one of the earliest clear statements of the free-rider problem, not least by Cornes and Sandler (1996) in their comprehensive treatment of the subject. However, given that Hume did not employ formal mathematical methods the exact nature of the public goods concerned is not entirely clear. This is particularly true of "justice". One objective of this paper, therefore, is to offer a simple framework to help clarify the exact public good nature of "justice". This will enable us to address the second question that we pose: what happens to the provision of "justice" as community size increases? As is well-known, Mancur Olson (1971) in his classic *Logic of Collective Action* hypothesised that larger communities would have more difficulty both in providing public goods through voluntary means and in achieving (Pareto) optimal provision of the same. The strong version of this conjecture is that provision would fall as the size of the community increases, and its weak version is simply that provision would move further from the optimum. As we shall see, Hume, in a rather basic way, makes parallel conjectures, although he does not discuss the question in any great detail or scope. In consequence, as Dougherty (2003) points out, it is misleading to think of Hume as having anticipated Olson's analysis and thus the modern development of the idea of a public good. Nevertheless, as Hardin emphasises, the prediction itself is central to Hume's thinking, a distinctive feature of which is that moral and political questions are analysed within the same framework. That the public good in question will not be satisfactorily provided in a large community provides the basis for a justification for the existence of the state.

Hume does not provide any extensive analytic justification for why "justice" will not be satisfactorily provided without the state, and it seems that commentators are for the most part happy with the fact that what he says is consistent with Olson's more detailed investigations. However, modern economic analysis of public goods, while broadly supportive of Olson's conjectures, does suggest qualifications. For example, with the standard (summation) form of public good Andreoni (1988) shows that the voluntary provision of the public good rises with the size of community (contrary to the strong version of the conjecture), and Hirshleifer (1983) points out that public good provision depends critically on the nature of the good itself.[1] Hence, before dealing with the question of community size we need a clear understanding of how "justice" is provided. To this end Sect. 1 of our paper

[1]As well as Cornes and Sandler (1996), Pecorino (2015) provides a good survey of this aspect of the economic theoretical literature stemming from Olson's work.

provides a brief exposition of Hume's theory of justice, as given in Volume Three of the *Treatise*. This goes over familiar ground, but its purpose is to motivate the formalisation to follow. Our model is set out in Sect. 2, which also explains, briefly, its relation to the existing technical literature on public goods. Section 3 shows how the (voluntary) provision of "justice" might depend on the size of the community. Concluding comments appear in Sect. 4.

1 Hume's Justice as a Public Good

In Volume 3 of the *Treatise* Hume discusses "justice", which he conceives of as a "convention enter'd into by all members of the society to bestow stability on the possession of .... external goods [ordinary commodities], and leave everyone in the peaceable enjoyment of what he may acquire by his fortune and industry" (*T*.3.2.2.489[2]). Individuals, in accepting such a convention, "regulate their conduct by certain rules", through which all parties gain. The analogy is drawn with the use of a single language, and with the emergence of "gold and silver as the common measures of exchange" (*T*.3.2.2.490). Hume makes a sharp distinction between "natural" and "artificial" virtues. The first of these includes a proper regard for the interest of oneself and one's family, friends and close acquaintances. Hume accepts that individual self-interest is the main motive force for human action, but he also allows for benevolence to others within one's personal sphere. "Artificial" virtues are those which further or at least are consistent with the maintenance of justice, and which induce acts which benefit those beyond our immediate personal sphere. It is likely indeed that no group of individuals is or ever was entirely without some conception of justice: "....every parent, in order to preserve peace among his children, must establish it [rule for the stability of possession]" (*T*.3.2.2.493). Individuals are assumed to be aware of the benefits that stem from justice, and once it has been established "every single [just] act is perform'd in expectation that others are to perform the like." (*T*.3.2.2.498).

Justice, in the *Treatise*, focuses on property and involves three key features:

(a) Stability of possession
(b) Transfer of ownership by consent
(c) The keeping of promises

The last of these is an extension of (b), given that important forms of exchange may require delivery at different points in time. These three concerns are discussed in sequence and within any community they evolve spontaneously in roughly the order given. The exact historical process by which this occurs is not spelt out, and is in any case likely to vary considerably from time to place. However, as

[2]References are to the Hume (1978), and indicate the Treatise (T), book part, section and page. References to Hume's Enquiry Concerning the Principles of Morals (1975) (EPM) give the section number.

Hume clearly states, there is a problem. I may well benefit from the existence of justice in some abstract sense but would I conform to any just convention in making actual decisions? He spends a lot of time analysing this question and, simplifying somewhat, comes up with three key answers as to why I would.

The first is Hobbesian in flavour. The consequences of a breakdown in justice are serious:

> "…. without justice, society must immediately dissolve, and every one must fall into that savage and solitary condition, which is infinitely worse than the worst situation that can possibly be suppos'd in society" (*T.*3.2.2.497).

But people will only conform if they expect others to do so as well, and breaches of justice will tend to be copied:

> "Your [bad] example….affords me a new reason for any breach of equity, by shewing me, that I should be the cully of my integrity, if I alone should impose on myself a severe restraint amidst the licentiousness of others." (*T.*3.2.7.535)

Self-interest therefore is the first motive for behaving justly. I behave honestly because I realise that otherwise, by tempting others to behave badly, I will induce others to act similarly and thereby bring adverse consequences on myself.

A second reason for adherence is reputation. In discussing promises Hume writes:

> "….I learn to do a service to another, without bearing him any real kindness; because I forsee that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or with others. And accordingly, after I have served him, and he is in possession of the advantage arising from my action, he is induced to perform his part; as forseeing the consequences of his refusal." (*T.*3.2.5.521)

These consequences are that "[he] must never expect to be trusted any more if he refuse to perform what he promis'd" (*T.*3.2.5.522).[3]

This motive is a variant of the first, but the adverse consequences of a failure to keep a promise are felt by me, and not necessarily by anyone else.

Both motives are consistent with self-interest in a narrow sense. To take an example, whether to drive on the left in the UK or the right in the USA is a social convention from which all benefit. We do not, however, consider those who adhere to this convention (more or less everyone) especially virtuous. The third motive introduces the idea of virtue, which is of particular importance in large communities where human interactions are more anonymous. Now private considerations (which include a regard for one's immediate circle) may induce individuals not always to act in accordance with "justice". Hume writes: "… when society has become numerous, and has encreas'd to a tribe or nation, this interest

[3]Bruni and Sugden (2000) discuss the nature of trust in the writing of Hume, Smith and Genovesi. Here by contrast, the public good nature of the actions considered, with no possibility for sanctions against individuals, means that trust and reputation play no direct role. The motive for performing the acts we consider is to preserve the social convention itself. Our first quotations suggest this was the important consideration for Hume. It also appears in the quotation they give on pages 30 to 31.

[justice] is more remote; nor do men so readily perceive, that disorder and confusion follow upon every breach of these rules, as in a more narrow and contracted society". (*T.*3.2.2.499)

This last observation leads to a discussion of sympathy and virtue. People may be self-interested, but they know when they have been harmed and, more importantly, through their capacity for sympathy, they recognise when others have been harmed. Once some form of convention has been established, even solely on the basis of self-interest, people appreciate the desirability of its continuation. Given that virtue and vice reflect the benefit or harm people experience, individuals develop a system of morality and apply it in their own behaviour. In this way justice, if established, changes the way people behave. Morality is in part a reminder of the importance of preserving desirable social conventions. Hume summarises:

> "*Thus self-interest is the original motive to the* establishment *of justice: but a* sympathy *with public interest is the source of the* moral approbation *which attends that virtue.*" (*T.*3.2.3.499) (Italics in the original)

Numbers therefore pose a threat to justice, and morality is a guard against this threat. However, whether it is sufficient by itself to ensure the stability of desirable rules of conduct is an open question. Hume seems to think not, as he envisages a role for government in bolstering morality through education. He claims that "moralists or politicians" can be expected "....to give a new direction to those natural passions, and teach us that we can better satisfy our appetites in an oblique and artificial manner, ..." (*T.*3.2.5.521). Simply exhorting individuals to pursue the public interest in their every day actions is unlikely to achieve much. Governments may recommend suitable behaviour, but this is likely to be pointless unless the implied conventions are such that all can expect to benefit (see *T.*3.2.2.500). The conjecture about numbers is therefore quite central to Hume's thinking in that it provides the foundation for his first argument for a role for the state.

Hume is very clear about the type of action individuals must undertake. They will be *discrete*. One should keep to contracts and promises. There are also non-actions that must be observed, such as respecting the property of others. In each case they must be performed by individuals. Hume writes (*T.*3.2.6.530) that "*property, and right and obligation admit not of degrees*....An object must either be in the possession of one person or another. An action must either be perform'd or not. The necessity there is of choosing one side in these dilemmas... oblige us ... to acknowledge, that all property and obligations are entire." [emphasis in the original].[4] Justice, in Hume's terms, is thus a convention from which all benefit. As such it is a public good which is provided and maintained over time by a sequence of discrete separate individual acts of all persons in the community. If individuals fail to perform the required acts then the provision of justice is undermined with possibly disastrous consequences.

[4]Dawkins (2003, pp. 24–25) comments on the discrete nature of lawyers' minds. Here is an explanation! Hume emphasises the often arbitrary nature of the judgements that will need to be made in implementing "justice".

Hume discusses a second type of public good, but spends little time on it. It appears at the end of the *Treatise* 3.2.7, and is clearly distinguished from justice. On concluding his discussion of the government's role in providing justice he writes: "But government extends further its beneficial influence; and not contented to protect men in those conventions they make for their mutual interest, it obliges them to make such conventions, and forces them to seek their own advantage, by a concurrence in some common end or purpose." (*T.*3.2.7.538). This introduces the conventional public goods much discussed by economists. As we know from the modern theoretical literature (see again Cornes and Sandler (1996)) they take many forms. It is the mode of provision that distinguishes them from Hume's "justice". Rather than being delivered over time by a series of discrete acts by separate individuals they require co-ordinated effort on the part of a (sub-)group of individuals within one period. As Hume famously writes:

> "Two neighbours may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each other's mind; and each must perceive, that the immediate consequence of his failing in his part, is, the abandoning the whole project. But 'tis very difficult, and indeed impossible, that a thousand persons shou'd agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expence, and wou'd lay the whole burden on others." (*T.*3.2.7.538).

This describes "market failure" resulting from a failure to *co-ordinate* individual action for a given purpose *in a single time period*. As stated in the introduction, this quotation is often used to illustrate the free-rider problem. This may, though, be misleading. Hume seems to be saying that the key reason for the failure of voluntary action lies in transaction (bargaining) costs which are increased as a result of the free-rider problem. In contrast to the 'justice' case he does not refer to a failure of individual perceptions. Hence again, albeit for a different reason, the presumption is that in a large community the outcome may not be fully satisfactory. We have therefore a second, and separate, role for the state:

> "Political society easily remedies both [agreement on and finance of a project] these inconveniences. Magistrates find an immediate interest in the interest of any considerable part of their subjects. They need consult no body but themselves to form any scheme for the promoting of that interest. And as the failure of any one piece in the execution is connected, tho' not immediately, with the failure of the whole, they prevent that failure, because they find no interest in it, either immediate or remote. Thus bridges are built; harbours open'd; ramparts rais'd; canals form'd; fleets equip'd; and armies disciplin'd every where, by the care of government, which, tho' compos'd of men subject to all human infirmities, becomes, by one of the finest and most subtle inventions imaginable, a composition, which is, in some measure, exempted from all these infirmities." (*T.*3.2.7.538)[5]

[5]Those sceptical of the ability of the state to provide goods and services efficiently will probably smile at this example of apparent naivety! However, Hume does not develop a concept of optimality. It is likely that what he actually had in mind was something very different: institutions can be developed under which self-interested individuals really do provide collective benefits. The very familiarity of this feature of modern democracies prevents us from seeing that perhaps it is a rather "subtle invention".

To summarise, there are two types of public good in Hume. Firstly, there is "justice". This requires discrete actions over time on the part of individuals in particular situations: keeping promises, respecting the property of others and so on. These are acts individuals must do themselves. No direct co-ordination of individual action is required, beyond the recognition that as a convention it will ultimately be undermined by individual failure to do what is required. The second type of public good corresponds to the standard public good typically analysed in the economic literature, and requires some form of direct co-ordination of individual action ("a concurrence in some common end or purpose"). This has no direct connection to justice as such, and Hume has comparatively little to say about it.

Many might take issue with Hume's notion of justice, with its exclusive concentration on property. This is a minor point, and may involve nothing more than semantics. His thinking easily extends to wider conceptions of justice. Not all promises, after all, concern property, and he himself points to the analogy with language or the use of precious metals as a medium of exchange. The same general reasoning applies to any social convention from which individuals derive mutual benefit. Indeed, the later *Enquiry Concerning the Principles of Morals* discusses "virtue" which has a broader meaning than the justice of the *Treatise* but is treated in a similar way (note the quotation on roads in the next section). In this paper we take the broader view, and use the terms "justice" and "social convention" synonymously (mostly the latter as it is the modern term). A second problem lies in the grey area of individual action and the law. To what extent can social conventions be upheld without legal sanction? Hume's view seems to be that in a large community individual action can take us some way, but it needs reinforcement by the government. This accords to some extent with some findings from modern economics. It is not too difficult to generate co-operation in a repeated prisoners' dilemma, and this topic is now a standard one in game theory texts. Axelrod (1984) is a classic early reference. Price collusion by firms, for example, is an illegal social convention but it is not especially uncommon.

The sharp division Hume makes between the two types of public good is reflected in current research in economic theory. The literature on "public goods" has focused mainly on the second ("co-ordination") type of good, while the social convention issue is addressed separately by game theorists, in particular with the Folk Theorem and its extensions. Binmore (2011) stresses how closely Hume anticipated the key problems posed by the latter and the notion of reciprocity that lies at its centre. To proceed, we note, following Hirshleifer (1983), that with public goods the efficiency of individual action depends on the specific way in which individual contributions aggregate to determine final provision. Given that Hume sets down very clearly a particular type of contribution structure, we proceed to analyse his conjecture on community size in a simple context using his general approach and adapting arguments that he himself used. Our argument is less an exposition of Hume than an enquiry into the voluntary provision of justice when community size is large. Nevertheless, we try to model certain key aspects of his thinking emphasised in this section. Hence we present, in a manner parallel to Sturn's view of the role of utility in Hume's thinking, a thought experiment to suggest how we might understand his ideas in the light of modern thinking.[6]

2 Modelling Justice

To proceed, three distinct features of his analysis need to be borne in mind:

(a) Individual acts sustaining justice are discrete;
(b) They are performed at different points in time by different individuals;
(c) They do not generally involve direct co-ordination with acts performed at the same time by other individuals.

To clarify the type of acts we have in mind it will be useful to start with Hardin's account of Hume's "strategic analysis". Individual interactions take three forms: conflict; co-operation; and co-ordination. In the political sphere the first of these, taking the form of distributive justice, plays no significant role in Hume's thinking and can be ignored.[7]

Neither will we here be much concerned with the problem of co-ordination. In the formal analysis to follow we assume there is some, possibly small, cost to performing the acts referred to above. Hume (*T.*3.2.2.497) refers to the possibility that an individual who performs a just act may thereby lose out in the current period. This is not true of the typical co-ordination problem, such as whether to drive on the right or the left. Whatever the convention is I clearly gain by conforming![8]

The category of interest in this paper therefore is co-operation. This is usually formalised as a prisoner's dilemma. There are two levels at which such a game might be played in Hume's account. First, all individuals might agree that adherence to some social convention is mutually beneficial. This situation is usually characterised as a prisoner's dilemma. We all gain if everyone adheres to some convention, although each of us would prefer others to do so, whilst we do not. If no one keeps to the convention then everyone is worse off as compared to universal adherence.

[6]To some small extent our model might possibly help explain why Hume, somewhat inconsistently, can be thought of as both a proto-game theorist and a proto-behavioural economist. The choice agents face is binary, so intransitivity is not a problem, and if we suppose that the situations we analyse are sufficiently similar so as to induce similar states of mind then the considerations discussed at length by Sugden (2006) are less likely to apply. Hume's use of "game theory" examples to make general points about how societies work seems in line with this procedure.

[7]Yellin (2000) points out that in his *Political Essays* Hume does admit that there may sometimes be a case for re-distribution (but not equality).

[8]Quite clearly, one would want to co-ordinate in this sense whatever number of agents one interacts with, and in the case of roads the individual benefit from so doing almost certainly increases with the size of the community. Hardin (2007, p. 94) points out that the legal requirement on which side of the road to drive seems in many cases to have arisen spontaneously.

Hume though rejected social contract theory: individuals do not decide formally to accept a social convention. Rather, they adhere to a convention when making actual choices. This involves a potential time consistency problem. I may agree that it is desirable in some abstract sense that everyone (including myself) should follow some social convention, but would I myself really behave in the desired way when confronted with an actual decision? Hume was aware of the difficulty. Indeed he gives us a very clear definition of the problem:

> "In reflecting on any action, which I am to perform a twelve-month hence, I always resolve to prefer the greater good . . . .. But on my nearer approach, those circumstances, which I first over-looked, begin to appear, and have an influence on my conduct and affections. A new inclination to the present good springs up, and makes it difficult for me to adhere inflexibly to my first purpose and resolution" (*T* 3.2.7.536).

As we have seen, his solution to this problem is bring in the government which can influence behaviour through education and the law. This last quotation comes in fact from the section entitled "Of the origin of government". Hume's concern in the *Treatise* is primarily a failure of perception. However, at the very end of the later *Enquiry Concerning the Principles of Morals*, without referring back to the analysis of the *Treatise*, he brings up the problem of the "sensible knave". This character is familiar to economic theorists as a genuine free-rider. He or she surreptitiously breaks the rules for his/her own interest, having made a calculation that he/she will not be caught. By contrast with the discussion in the *Treatise* the problem is now one of too much perception! The key issue is this: is it the case, particularly in a large community, that people would want to adhere to the convention, or will we need some form of coercion or inducement from the state to ensure that they do? [9]

The second level of the game (the stage game) may or may not involve a prisoner's dilemma. If so, Hume's problem corresponds to that addressed by the literature on the Folk Theorem, in which individuals play some form of repeated prisoner's dilemma (Fudenberg and Maskin 1986). Here, identified individuals

[9]This question has caused debate amongst philosophers, and is related to the question of how we treat the "true interest" of the sensible knave. According to Baron (1982) and Gauthier (1992) Hume has an "error" theory of virtue in that it is *not* in the interest of all (notably the sensible knave) to be virtuous, although "moralists and politicians" must inculcate the view that it is. Baier (1992) counters by employing Hume's ideas on the evolution of moral thinking as societies develop. Culp (2013) surveys the issue, concluding that that Hume was wrong if he really argued that adherence to justice is in the interest of literally everyone, but that he probably didn't think this himself. (The final section of the *Enquiry*, at the end of which the sensible knave appears, aims to "inquire whether every man . . . .will not best find his account in the practice of every moral duty". The answer to this question is not explicitly given.) His reluctance to spell matters out probably came from the danger of undermining adherence to justice itself. However, if the sensible knave is a sufficiently small exception to the generality of mankind, education for justice is indeed in the true interest of (nearly) all and in that sense an honest policy. It would surely make little difference to Hume's thinking in the *Treatise* to allow for the fact that the application of the law provides incentives even for a sensible knave to conform to conventions and possibly even thereby develop the moral qualities of which he approves. We are dealing with a *public* good: free-riding sensible knaves benefit even if they do not properly contribute.

play a prisoner's dilemma at each stage and, in line with Hume's comments on the keeping of promises quoted earlier, the (credible) threat of deviators being individually punished ensures that the social convention of co-operation is adhered to. In many cases of public interaction, though, the acts to be performed are done anonymously, that is without any expectation that the actor (or non-actor) will be observed by the same people on a second occasion. Hume says little about this situation, except for a few thoughts concerning the "sensible knave", but it is clearly important. Within the Folk Theorem literature, assuming two person interaction at each stage, the implications of anonymity have been analysed by Ellison (1994) building on previous work by Kandori (1992). It seems from this work that a social convention to co-operate can be stable in a large community although perhaps not when community size is very large. Hume's thinking, however, clearly goes well beyond situations in which the stage game is a prisoner's dilemma. Xie and Lee (2012), for example, analyse a repeated game involving trust (with anonymity), with similar results. There is, though, another possibility which seems not to have received so much formal attention.

This is the case of "waggoners" given by Hume in the later *Enquiry*: "Waggoners, coachmen, and postilions have principles, by which they give the way; and these are chiefly founded on mutual ease and convenience." (*EPM* 4). This has a modern counterpart in the common practice of British drivers already on a motorway or dual carriageway, to pull over into the (outside) lane to the right, when this is safe, so as to allow those entering easy access. This is in spite of the fact that they have the right of way. Such acts of courtesy have induced some (Cassini 2006) to question whether the widespread use of traffic control measures is desirable. The spontaneous conventions that motorists would develop in their absence could ensure safer roads and a freer flow of traffic. With this type of situation only one person is able to perform the required act. There is therefore no proper stage game in the set-up. Also note the slight inconvenience incurred by waggoners when they "give the way": they lose out in the current period. A parallel example may be familiar to the reader, that of refereeing a paper for a journal. Here again just one person can perform the act (let us say), but now there is a direct benefit to many individuals (academic standards in one's chosen field), rather than to just one.

To see whether a social convention of giving the way is viable we model the economy as follows. Ex ante, each period is the same except for the preceding history. At the beginning of each period nature makes two moves. First, from the population of n individuals she selects one who is placed in a position to do some act, as perhaps outlined in our examples. The second move is to determine the cost of performing the act in question. This random variable is denoted by $x \in [L, U]$, and its realisation by c_i, i indicating the selected individual. This reflects the fact that circumstances vary as between periods. The ease of letting someone onto the motorway will vary according to the state of traffic, and the cost to any potential referee will vary according to the nature of the paper. Both forms of randomisation are assumed identical across individuals and periods. We assume a well-behaved distribution function $F(x)$ which is common knowledge to all agents as is its realisation c_i. We assume an infinite number of periods.

The public good we focus on here is the existence and sustainability of the social convention itself. This requires an individual who is in a position to perform the required act in any period to do so as long as their realised cost is not too high. Formally, this action must be done whenever $c_i < \overline{c}$ where $\overline{c}$ is some maximum cost level which has perhaps evolved through practice. The benefit to each individual is therefore the expected present value of having the convention, and individuals 'contribute' by performing acts in accordance with this convention. Both types of problem we treat have been discussed previously. Bergstrom (2012, 2013) has discussed the "Good Samaritan" problem in a one period context, focusing on the decision on whether to help an individual in distress when others can do so and when people do not observe that individual at the same time. Individuals are motivated by benevolence and the key question is whether a large number of potential helpers results in the victim having to wait longer to receive help. The refereeing problem is a discrete version of Hirshleifer's best-shot public good.[10] In what follows we assume the identity of the focal individual in each period to be common knowledge. In consequence, the social convention requires a specified individual each period to act in a particular way.

3 Adhering to a Social Convention

According to Hume the key motive for performing acts of justice lies in the preservation of 'justice' itself. One might be prepared to perform a costly task today if in return one expected to gain from the existence of the convention in some future period. Before any period starts there will be an expected ex ante benefit per person should the social convention be adhered to. This might simply be the anticipated benefit of having some public good provided (refereeing), in which case we denote this benefit by b. Alternatively (waggoners) the benefit might represent the expected benefit to a single individual of having a Good Samaritan. The cases differ slightly. We deal with each in turn.

3.1 Refereeing a Paper

Assume in each period each person in a community of n has an equal chance of having to perform a task. The probability that they will be called to act is $\frac{1}{n}$. For any future period the ex ante net benefit *to each individual* if the task is performed is

[10]It is not too difficult to extend our analysis to the case in which a subset of the community is in a position to perform the act. This encompasses the bystander problem of social psychology. The conclusions in this case are strengthened by the fact that with a larger group of potential helpers the expected minimum cost is lower and the task more likely to be performed.

$\left[b - \frac{1}{n}E\left(c\middle|c < \overline{c}\right)\right]$. We shall assume $b < L$, so that the focal individual is bound to lose in the current period should he/she perform the desired act. However, the ex ante benefit is assumed to be positive. Under the social convention he/she is expected to perform the task, but only if the cost is not too high. The social convention is only under threat when agents with costs less than $\overline{c}$ do not perform the action. Given this social convention the ex ante expected gain per period to each individual is:

$$F\left(\overline{c}\right)\left[b - \frac{1}{n}E\left(c\middle|c < \overline{c}\right)\right]$$

The present value (ex ante) of having the social convention would therefore be:

$$\frac{1}{1-\delta}F_m\left(\overline{c}\right)\left[b - \frac{1}{n}E\left(c\middle|c < \overline{c}\right)\right]$$

δ is the discount factor. This expression would represent the ex ante gain to each individual from mutual co-operation when there is a social convention in operation (level 1 prisoner's dilemma as explained earlier). Now Hume poses the time consistency problem. If this expression is positive the social convention might be desirable in some social contract sense, but would the focal individual actually want to act in accordance with the convention?

As is well-known, individual behaviour for this sort of problem can modelled in various ways. Here we assume a simple strategy:

1. *In the first period, if called to act, do so if and only if the cost c_i is less than the critical value $\overline{c}$.*
2. *In subsequent periods, if in a position to act, do so if: (a) the cost c_i is less than the critical value $\overline{c}$; and (b) in all previous periods focal individuals with a cost less than the critical level $\overline{c}$ acted. Otherwise do not act.*

Hence agents adopt a version of the grim or trigger strategy. Two arguments justify it, beyond simplicity. Firstly, Hume himself hints at something along these lines when he says "....disorder and confusion follow upon *every* breach of these rules ...." (emphasis added). Secondly, grim is the most unforgiving of co-operative strategies, so the incentive to maintain a co-operative outcome is as strong as it can be. If the convention cannot be sustained with the trigger strategy, it will not be possible under any other strategy. In this way our analysis provides necessary conditions for a social convention to be sustainable. Clearly in practice more "realistic" strategies can be devised, ones that allow perhaps for mistakes. In line with Hume's own thinking about habit and inertia in human behaviour people might decide whether to adhere to the convention only at periodical intervals, and will do so as long as previous violations of the convention have not been too great. It is not clear that this approach would alter our fundamental conclusions, so for present purposes we keep to our simple form of grim. The key idea, from Hume, is that people will perform acts that benefit others, but only when they expect others to adhere to the social convention involved. Admittedly grim weights the conclusions in favour of adherence, but bear in mind that we assume there is no way of punishing just the person who does not adhere to the social convention. This points in the opposite direction.[11,12]

In each period when the social convention deems that the task should be done the focal individual performs the task when the following condition holds:

$$G(\overline{c}, n) \equiv b + \frac{\delta}{1-\delta} F(\overline{c}) \left[b - \frac{1}{n} E\left(c \middle| c < \overline{c} \right) \right] \geq c_i \tag{1}$$

The left-hand side of this inequality is the present value of the social convention (public good). The public good nature of justice manifests itself at this point, as b (or something similar in a heterogeneous community) also represents the benefit to other members of the community.

To see whether any value $\overline{c}$ could sustain a sub-game perfect equilibrium substitute $\overline{c}$ for c_i. If the inequality is preserved a sub-game perfect equilibrium adhering to the social convention with $\overline{c}$ is possible: there is some $\overline{c} \in (L, U]$ such that:

$$G(\overline{c}, n) \equiv b + \frac{\delta}{1-\delta} F(\overline{c}) \left[b - \frac{1}{n} E\left(c \middle| c < \overline{c} \right) \right] \geq \overline{c} \tag{2a}$$

It is obvious that G rises with n, with the limit being $b + \frac{\delta}{1-\delta} F(\overline{c})\, b$ as n goes to infinity.

If $\overline{c}$ is above this limit then the proposed social convention is not viable, in the sense that with probability one the convention will break down in finite time: agents, noticing this, will not adhere to the convention in any period. This point has some bearing on the question of universal adherence mentioned in footnote 9, as $\overline{c}$ must be sufficiently low for the social convention to be viable, and the lower it is the more likely it is to be sustainable. For a sufficiently high discount factor any given $\overline{c}$ is viable. Below the limit, however, "co-operation" will always be possible for n sufficiently large. Note too (formally) that if $\overline{c} > L > b$ and $n = 1$ the expression in square brackets is negative so that $G(\overline{c}, n) < \overline{c}$ is possible: we may need a sufficient number of people in the community for the convention to be viable.

[11]Both Kandori and Ellison produce Folk Theorems for cases where, in a prisoner's dilemma context, punishment strategies are less harsh than simple grim, but where, at each stage, play is anonymous. The implications of community size are less clear than in the circumstances we analyse here.

[12]In the literature on organisational behaviour there is evidence indicating that the efficiency of teams is adversely affected by commitment shown by the least co-operative individual of the team. See Raver et al. (2012) and the references contained therein.

3.2 Waggoners

Here we assume that in each period there is a probability p that any individual will require help of some sort. This can only happen once in any period to any one individual, but we assume any number of individuals could require help in any one period. The benefit of being helped we denote by B. Whenever help is required the nature of the incident is such that only one person is in a position to provide help, as with Hume's waggoners. Hence, assuming the social convention to be in place, the expected net benefit in dealing with one incident in any future period is:

$$\left[pB + b - pE\left(c\middle|c < \overline{c}\right)\right] \tag{3}$$

To understand this expression note that the expected number of incidents in which any specified individual does not require help is $(n-1)p$, and the probability of that individual being focal for any one incident is $(n-1)^{-1}$. b now represents a feeling of sympathy from living in a society in which individuals help one another, and might reflect the moral sense that Hume suggests will develop particularly in a larger community. Most probably this term would take the value zero were our individual to be a sensible knave. Otherwise it could be multiplied by $(n-1)p$ if this benefit adheres to each incident in which help is given. Both interpretations are consistent with what Hume wrote about sympathy, but it is also possible this term should not be here at all.[13] Condition (2a) now becomes:

$$G\left(\overline{c}, n\right) \equiv b + \frac{\delta}{1-\delta} F\left(\overline{c}\right)\left[pB + b - pE\left(c\middle|c < \overline{c}\right)\right] \geq \overline{c} \tag{2b}$$

This inequality is invariant with respect to n, with the possibility that it is more likely to hold if we suppose sympathy to be part of individuals' make-up, and that this adheres to the observation of individual acts. In this case our conclusions are as before, but however we read b, a high value of n does not threaten the social convention. The G function in each case represents the advantages of living in a society in which a social convention is adhered to. Apart from the b term in inequality (2b) it would take the same form whether the individual was a sensible knave or not. Regardless of the individual's type a large community does not seem inimical to the continuation of the convention itself. We discuss the implications of these results for Hume's thinking in the conclusion.

[13] See Appendix 2 of the *Enquiry* or the essay "On the dignity or meanness of human nature" Hume (1987). Garrett (2007, p. 270) lists a series of points made by Hume suggesting non-egotistical reasons for adhering to a social convention.

4 Conclusion

In this paper we have tried to document the relationship of Hume's "justice" to the idea of a public good, and then to formalise the idea of a social convention using ideas he clearly expressed himself. For the former task we find much in the *Treatise* that corresponds to modern thinking, with, perhaps rather strangely, the possible exception of the free-rider problem. This indeed has only a rather secondary role in co-ordination types of public goods, and is only addressed in the "justice" case in the later *Enquiry*, where he was reluctant to admit its importance. For the latter problem our formalisation suggested that any convention ($\overline{c}$) attainable in a small community is also attainable in a large community. The explanation for this seems to result from the inter-temporal trade aspect of the model, where "every single [just] act is perform'd in expectation that others are to perform the like." In one way this finding is not surprising: that (sustainable) social conventions differ across nations is something which strikes even the most unobservant of travellers. However, there remain problems. As with earlier work on the Folk Theorem, the grim strategy seems implausible in many social contexts, as does the assumption of common knowledge (though as we are dealing with some form of public good the assumption of complete ignorance is even less plausible). Hume writes that men do not ". . . .so readily perceive, that disorder and confusion follow upon *every* breach of these rules, . . . " (emphasis added). Most would probably agree with the first part of the quotation, but might question the last clause.

It is therefore worth pausing for a moment to think further about the problems that any social convention itself might face. A key feature of our model is that there is a focal individual who is obliged by the social convention to perform the desired act. This person is observed by everyone, as is their cost of performing the act. Everyone understands what has to be done and who is to do it. However, in practice actions may not be observed immediately, and even if they were individuals may not be able to assess whether the social convention had been adhered to, or indeed what precisely it requires. Even benign individuals might have difficulties is working out what to do in many situations. These are considerations which might, on Hume's argument, argue for a role for the state even with relatively small communities.[14]

A game of incomplete information might therefore be a more sophisticated way of capturing some of Hume's insights, but conventional game theory itself does not fully capture what he was saying. Recall his dynamic view of human beings in society. Self-interest is the first motive for behaving justly, and is later strengthened through developing notions of virtue and morality. A collection of

[14]Our conclusions extend, without much difficulty to the bystander problem of social psychology. As observed in one recent textbook (Hewstone et al. 2005, p. 387): "Numerous studies indicate that the willingness to intervene in emergencies is higher when a bystander is alone." Explanation for this failure to act runs along three different lines: diffusion of responsibility; ignorance of how others interpret the event; and concern about how one's own behaviour will be interpreted. These considerations all reflect confusion about the nature of the situation people find themselves in.

individuals could only be called a society if it were reasonably stable over time, with individuals expecting to interact in the future. The G function of equations (2) in the first instance reflected pure self-interest resulting from future interactions. If community size is small and interactions fairly personal this might be enough for social conventions to become established. On Hume's account a large community brings problems stemming from a failure of perception. This could perhaps be modelled as a fall in the discount factor or a lowering of expected future benefit as n increases. However, if moralists and politicians succeed in their task of informing individuals about the dangers of a collapse in social norms the equations (2) would seem to suggest no extra need for morality in a large community. In line with Hume's thinking in the *Treatise*, large numbers are not per se a problem. It is rather the information problems that they pose that causes the difficulty. A sense of morality would bolster people's willingness to conform to the social conventions.

So our analysis actually confirms one insight implicit in the *Treatise. If* people are accurately and fully informed (however that may be) about the consequences of deviating from a social convention community size has no special implications for the viability of the conventions themselves. This statement also holds for sensible knaves. These people may or may not conform depending on circumstances, but these circumstances do not necessarily include community size. Of course, a larger community is likely to mean greater anonymity with more opportunities for knavery. Hence, as indicated in footnote 9, problems remain if we want to ensure that *all* individuals willingly adhere to social norms. It seems difficult to believe that in practice in a heterogeneous community the inequalities (2) could hold for all, even with a widespread sense of morality. However, if we agree that sensible knaves are a small minority there would be few problems for the overall structure of Hume's thought. Little is surely conceded by extending government's role beyond education or information provision to that of changing the incentives facing sensible knaves.

References

Andreoni, J. (1988). Privately provided goods in a large economy: The limits of altruism. *Journal of Public Economics, 35*, 57–73.

Axelrod, R. (1984). *The evolution of co-operation*. New York: Basic Books.

Baier, A. (1992). Artificial virtues and the equally sensible non-knave: A response to Gauthier. *Hume Studies, 18*(2), 429–440.

Baron, M. (1982). Hume's noble lie: An account of his artificial virtues. *Canadian Journal of Philosophy, 12*(3), 539–555.

Bergstrom, T. C. (2012). *Single-payer volunteers' dilemma*. Santa Barbara: University of California.

Bergstrom, T. C. (2013). *The Good Samaritan and traffic on the road to Jericho*. Santa Barbara: University of California.

Binmore, K. (2005). *Natural justice*. Oxford: Oxford University Press.

Binmore, K. (2011). *David Hume: Grandfather of modern economics*. www.sire.ac.uk/.../revised_SIRE_DavidHumeConference2011.pdf

Bruni, L., & Sugden, R. (2000). Moral canals: Trust and social capital in the work of Hume, Smith and Genovesi. *Economics and Philosophy, 16*, 21–45.
Cassini, M. (2006). In your car no one can hear you scream! Are traffic controls in cities a necessary evil? *Journal of Economic Affairs, 26*(4), 75–78.
Cornes, R., & Sandler, T. (1996). *The theory of externalities, public goods and club goods* (2nd ed.). Cambridge, UK: Cambridge University Press.
Culp, J. (2013). Justice, happiness, and the sensible knave: Hume's incomplete defense of the just life. *The Review of Politics, 75*, 193–219.
Dawkins, R. (2003). *A devil's chaplain*. London: Phoenix.
Dougherty, K. L. (2003). Public goods theory from eighteenth century political philosophy to twentieth century economics. *Public Choice, 117*, 239–253.
Dow, S. C. (2009). David Hume and modern economics. *Capitalism and Society, 4*(1), 1–29.
Ellison, G. (1994). Co-operation in the Prisoner's Dilemma with anonymous matching. *Review of Economics Studies, 61*, 567–588.
Fudenberg, D., & Maskin, F. (1986). The Folk theorem in repeated games with discounting or with incomplete information. *Econometrica, 54*, 533–554.
Garrett, D. (2007). The first motive to justice: Hume's circle argument squared. *Hume Papers, 33*(2), 257–288.
Gauthier, D. (1992). Artificial virtues and the sensible knave. *Hume Papers, 18*(2), 401–428.
Hardin, R. (2007). *David Hume: Moral and political theorist*. New York: Oxford University Press.
Hewstone, M., Fincham, F. D., & Foster, J. (2005). *Psychology*. Oxford: Blackwell.
Hirshleifer, J. (1983). From weakest-link to best-shot: The voluntary provision of public goods. *Public Choice, 41*, 371–386.
Hume, D. (1975). *An enquiry concerning the principles of morals* (3rd ed., edited by Lewis A. Selby-Bigge and Peter H. Nidditch). Oxford: Oxford University Press (SBN in the text followed by page number).
Hume, D. (1978). *A treatise on human nature* (2nd ed., edited by Lewis A. Selby-Bigge and Peter H. Nidditch). Oxford: Oxford University Press.
Hume, D. (1987). *Essays moral political and literary*. Indianapolis: Liberty Fund.
Kandori, M. (1992). Social norms and community enforcement. *Review of Economic Studies, 59*, 63–80.
Levy, D. (1984). David Hume's invisible hand in *The Wealth of Nations*: The public choice of moral information. *Hume Studies, 10*, 110–149.
Nuyen, A. T. (1986). Hume's justice as a collective good. *Hume Studies, 12*(1), 39–56.
Olson, M. (1971). *The logic of collective action*. Cambridge MA: Harvard University Press.
Pecorino, P. (2015). Olson's *logic of collective action*" at fifty. *Public Choice, 162*, 243–262.
Raver, J. L., Ehrhart, M. G., & Chadwick, I. C. (2012). The emergence of team helping norms: Foundations within members' attributes and behaviour. *Journal of Organizational Behavior, 33*, 616–637.
Sturn, R. (2004). The sceptic as an economist's philosopher: Humean utility as a positive principle. *The European Journal of the History of Economic Thought, 11*(3), 345–375.
Sugden, R. (2005). Why rationality is *not* a consequence of Hume's theory of choice. *The European Journal of the History of Economic Thought, 12*(1), 113–118.
Sugden, R. (2006). Hume's non-instrumental and non-propositional decision theory. *Economics and Philosophy, 22*(3), 365–391.
Vanderschraaf, P. (1998). The informal game theory in Hume's account of convention. *Economics and Philosophy, 14*(2), 215–247.
Xie, H., & Lee, Y.-J. (2012). Social norms and trust among strangers. *Games and Economic Behaviour, 76*(2), 548–555.
Yellin, M. E. (2000). Indirect utility, justice, and equality in the political thought of David Hume. *Critical Review, 14*(4), 375–389.

Decentralized Leadership

Emilson Caputo Delfino Silva

1 Introduction

In a family context, in which the well-known "rotten kid theorem" holds, benevolent parents are unable to commit to incentive schemes (carrots and sticks) to induce their selfish children to be well behaved [see, e.g., Becker (1981), Bergstrom (1989) and Cornes and Silva (1999)]. Children may choose actions to promote their self-interests in lieu of their family's common good. However, parents are the family workers and control the allocation of bequests, which occurs after parents observe their children's actions. Becker (1981) demonstrates that selfish children are well behaved (i.e., they maximize their family's welfare) if they anticipate that they are personally better off by taking actions that maximize family income. Bergstrom (1989) shows that Becker's rotten kid theorem is not general, but it holds if the children have quasilinear preferences and their wellbeing are normal goods for their parents. Cornes and Silva (1999) show that the rotten kid theorem holds when the kids' preferences can be represented by general but identical continuous and concave utility functions for two normal goods, a private good (numeraire) and a standard pure public good.

The interactions between selfish children and a loving, benevolent, parent in a family is similar in many respects to those between self-interested regional and benevolent central governments in federations where regional governments have (some) policy autonomy. In many federations, regional governments are able to implement some policies without seeking approval or support from federal authorities. The literature refers to this phenomenon as "decentralized leadership." In Canada, the provinces of British Columbia and Alberta have been leaders in

E.C.D. Silva (✉)
Department of Marketing, Business Economics and Law, University of Alberta, Edmonton, AB, Canada T6G 2R6
e-mail: emilson@ualberta.ca

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_9

global environmental policy—they unilaterally moved forward with the levying of carbon taxes even though a nationwide policy on carbon emissions is still lacking. In the European Union, the nation states have considerable power vis-à-vis the central government to adopt several types of policies, ranging from policies that govern the provision of local public goods (e.g., policing, health care, national security) to policies that determine their contributions to public goods that generate transnational benefits (e.g., abatement of carbon emissions). In such federations, characterized by decentralized leadership, one also observes substantial interregional income and fiscal transfers.[1] These reflect policies that central authorities control and which intend to reduce income and fiscal disparities across regions.

Caplan et al. (2000), motivated by the parallels between families and federations, demonstrate that (two) self-interested regional governments provide efficient contributions to a federal, pure public good if they make these contributions knowing that the benevolent central government will redistribute income across regions after it observes the regional governments' actions. The authors show that this rotten kid theorem continues to hold when one extends the model to allow imperfectly mobile residents to choose their region of residence after they observe the policies implemented by regional and central authorities. Residents are imperfectly mobile due to idiosyncratic regional attachment benefits (e.g., language, culture, customs, family ties).

This paper revisits the efficiency of decentralized leadership, the main issue studied by Caplan et al. (2000), but in situations in which regional governments provide multiple public goods and the central government controls multiple interregional-transfer instruments.[2] Each regional government contributes to the provision of two public goods, one of which is regional and the other is federal. The quantity of federal public good is an aggregation of regional contributions where the aggregation consumption technology is represented by a concave function. Particular cases of this concave function are summation (pure public good) and Cobb-Douglas (weaker link). Cornes (1993) and Cornes and Hartley (2007) advanced the study of weaker-link public goods. At the federal level, good examples are control of infectious diseases and counterterrorism effort.

The central government is responsible for interregional income and earmarked fiscal transfers. To facilitate comparisons and illustrate the social desirability of earmarked fiscal transfers, the center's policy arsenal is initially restricted to contain an instrument to implement interregional income transfers only. The initial setting is further restricted with the assumption that individuals are immobile.

[1]See, e.g., Silva (2015, 2017). These papers provide ample evidence of the importance of interregional earmarked and income transfers in several federations.

[2]See Cornes and Itaya (2010) for an interesting study of voluntary contributions to multiple pure public goods. The authors show that the provision levels in equilibrium are too low (relative to efficient levels), among other things.

The results of this paper demonstrate that there are efficiency-enhancing incentives promoted by centralized income and earmarked transfers when used together. In the simpler model without residential mobility, interregional income transfers equalize marginal utilities of income. Interregional earmarked transfers, implemented to reduce fiscal disparities in the provision of regional public goods, equalize marginal utilities of consumption of regional public goods. With separable (or homothetic) utility functions, the centralized choices imply equalization of consumption of private and regional public goods across regions, which lead forwarding looking regional governments to realize that they wish to maximize the same objective function. In the presence of such perfect incentive equivalence, regional governments provide efficient contributions to regional and federal public goods. In the absence of earmarked transfers, regional governments provide efficient contributions to the federal public good, but overprovide the regional public goods when they are able to commit to provision of both types of public goods. As in Caplan et al. (2000), the interregional income transfers promote incentives for efficient decentralized behavior on the provision of a federal public good. However, they also essentially make the regional governments to treat regional public goods as a federal public good, since the interregional-income-transfer mechanism creates a "universal" subsidy rate for the provision of all public goods that are subject to decentralized leadership. By including the earmarked fiscal transfer in the arsenal of instruments controlled by the center, one perfectly cures the anomaly caused by interregional income transfers, since the implicit subsidy rate disappears in the provision of regional public goods.

The efficiency of decentralized leadership remains in the extended model with imperfect residential mobility provided there is a common labor market in the federal economy. The existence of a common labor market implies that individual choices of region to work and to reside are independent. In such circumstances, interregional income and fiscal transfers lead to equalization of consumption of private and regional public goods across regions, and hence perfect incentive equivalence.

This paper contributes to multiple branches of the fiscal federalism literature.[3] It is the first paper, to my knowledge, to examine the efficiency of decentralized leadership in a setting in which regional governments commit to the provision of regional and federal public goods and the center is endowed with instruments to implement income and earmarked interregional transfers. Silva (2014) considers a decentralized leadership setting in which regional governments provide regional and federal public goods and the center promotes interregional transfers. However, the center does not have an instrument to implement earmarked transfers. Silva (2015) examines the efficiency of interregional earmarked fiscal transfers when used together with interregional income transfers in a federation where regional

[3]Please see Silva (2014, 2015) and Silva and Lucas (2016) for important contributions to fiscal federalism in the areas of decentralized leadership, earmarking and soft budgets, and imperfect residential mobility due to regional attachment, respectively.

governments provide multiple regional public goods only. This paper is also the first to combine the policy ingredients described above with a common labor market in the extended model with imperfectly mobile residents. Silva and Lucas (2016) show that decentralized leadership is efficient in the presence of a common labor market and imperfectly mobile residents when regional governments provide multiple types of federal public goods, but no regional public goods, and the center implements interregional income transfers. In the absence of regional public goods, the center does not need to have an instrument to implement earmarked transfers in order to promote incentives for efficient behavior at the regional level.

In what follows, Sect. 2 introduces the basic model, and Sect. 3 examines the socially optimal allocation and the subgame perfect equilibria for games in decentralized leadership settings. Individuals are assumed to be immobile in Sect. 3. In Sect. 4, the model is extended to allow residential mobility. In this section, one first considers the socially optimal allocation and later examines the subgame perfect equilibrium for a game with decentralized leadership where the regional governments provide contributions to both regional and federal public goods prior to the center's choices of interregional income and earmarked transfers. Section 5 offers conclusion remarks and suggestions for future research.

2 Basic Model

Consider an economy with two regions. There are n_i residents in region i, $i = 1, 2$, and $N = n_1 + n_2$ in the economy. There are two regional governments and one central government. Each region contributes to the provision of two types of public goods, regional and federal. Region i provides y_i units of the regional public good and g_i units of the federal public good. The contributions of both regions to the federal public good are aggregated according to an aggregation consumption technology, which is formally described by the function $f(g_1, g_2)$, where $f(.)$ is increasing in each argument and concave. The regional contributions to the federal public good produce Q units of this good; that is, $Q = f(g_1, g_2)$. Common examples of the federal public good are the pure public good, $Q = g_1 + g_2$ (e.g., national defense and abatement of greenhouse gases), and the weaker-link public good, $Q = g_1^{\alpha} g_2^{1-\alpha}$, $0 < \alpha < 1$ (e.g., control of contagious diseases and national border management).

The representative resident of region i derives utility $u(x_i, y_i, Q)$ from consumption of x_i units of a numeraire good, y_i units of the regional public good and Q units of the federal public good. For simplicity, the utility function is assumed to be strongly separable: $u(x_i, y_i, Q) = b(x_i) + r(y_i) + v(Q)$, where $b' > 0$, $b'' < 0$, $r' > 0$, $r'' < 0$, $v' > 0$, $v'' < 0$.

The provision of each type of public good has a unitary cost equal to one unit of the numeraire good. In the setting with immobile residents, the representative resident of region i is endowed with a fixed amount of income, w_i. In this case, region i's budget constraint is $n_i x_i + g_i + y_i = n_i w_i$ in the absence of transfers controlled by the central government. In the setting in which residents are mobile, the representative resident of region i earns market income from supplying labor and from profit shareholding.

The governments are utilitarian. In the setting with immobile residents, the payoffs for regional and central governments are $U^i = n_i [b(x_i) + r(y_i) + v(Q)]$, $i = 1, 2$, and $U = U^1 + U^2$, respectively. In the setting with mobile residents, the payoffs for regional and central governments also include psychic attachment benefits, to be described in detail in Sect. 4.

3 Economy with Immobile Residents

In this section, assume that $n_1 = n_2 = n = N/2$ and $w_1 > w_2$. These assumptions guarantee that region 1 is richer than region 2, which then provides the impetus for the central government to implement interregional income transfers. The assumption that the regions are equally populated is also consistent with all equilibria examined in Sect. 4. Therefore, it facilitates comparisons and helps the reader to understand the key rationale for the results.

3.1 *Social Optimum*

The social planner chooses non-negative $\{x_1, x_2, y_1, y_2, g_1, g_2\}$ to maximize

$$n[b(x_1) + b(x_2) + r(y_1) + r(y_2)] + Nv(Q) \tag{1}$$

subject to the national resource constraint:

$$n(x_1 + x_2) + G + Y = W, \tag{2}$$

where $G = g_1 + g_2$, $Y = y_1 + y_2$ and $W = n(w_1 + w_2)$. Letting $\lambda \geq 0$ denote the Lagrangian multiplier associated constraint (2), the first order conditions are as follows (for $i = 1, 2$):

$$nb'(x_i) - n\lambda = 0, \tag{3}$$

$$nr'(y_i) - \lambda = 0, \tag{4}$$

$$Nf_i v'(Q) - \lambda = 0. \tag{5}$$

Combining equations (3) and (4) yields

$$\frac{nr'(y_i)}{b'(x_i)} = 1. \tag{6}$$

Combining equations (3) and (5) yields

$$\frac{Nf_i v'(Q)}{b'(x_i)} = 1. \tag{7}$$

Equations (6) state that the regional public goods are provided at levels that equate each region's sum of marginal rates of substitution between regional and numeraire goods to the marginal cost of provision. Equations (7) show that the socially optimal contributions to the federal public good equate the nation's sum of marginal rates of substitution between each region's contribution to their marginal contribution costs. Equations (3) and (4) imply that $x_1 = x_2$ and $y_1 = y_2$, respectively, because $b'' < 0$ and $r'' < 0$.

3.2 Decentralized Leadership

3.2.1 Unlimited Regional Commitments

The regional governments are able to commit to contributions to regional and federal public goods. The center observes these contributions and then implements interregional income transfers. Let τ_i denote the amount of income transfer that region i receives (if positive) or pays (if negative). Region i's budget constraint is

$$nx_i + g_i + y_i = nw_i + \tau_i, \quad i = 1, 2. \tag{8}$$

The center's income transfers are redistributive. Hence,

$$\tau_1 + \tau_2 = 0. \tag{9}$$

The center chooses $\{\tau_1, \tau_2\}$ to maximize social welfare (1) subject to constraints (8) and (9). The first order conditions yield constraints (8), (9) and

$$b'(x_1) = b'(x_2). \tag{10}$$

Since $b'' < 0$, equation (10) implies that $x_1 = x_2 = x$. Combining this result with equations (8) and (9) yields $Nx + G + Y = W$. For simplicity, one can express the center's response function in terms of the numeraire good rather than in terms of the income transfer instruments. Hence, let $x(g_1, g_2, y_1, y_2)$ denote the center's best response function. The national resource constraint enables one to write

$$x(g_1, g_2, y_1, y_2) = \frac{W - G - Y}{N}. \tag{11}$$

In the first stage of the game, regional government i chooses non-negative $\{g_i, y_i\}$ to maximize $n\left[b\left(x\left(g_1, g_2, y_1, y_2\right)\right) + r\left(y_i\right) + v\left(f\left(g_1, g_2\right)\right)\right]$, taking the choices of the other region as given. The first order conditions yield (for $i = 1, 2$)

$$\frac{Nf_i v'(Q)}{b'\left(x_i\right)} = 1, \tag{12}$$

$$\frac{nr'\left(y_i\right)}{b'\left(x_i\right)} = \frac{1}{2}. \tag{13}$$

Conditions (12) are the modified Samuelson conditions for optimal contributions to the federal public good. Conditions (13) are the conditions that determine the contributions to the regional public goods. They equate each region's sum of the marginal rates of substitution between consumption of regional public good and numeraire good to the perceived marginal rate of transformation between regional and numeraire good. The latter is distorted by the income transfer mechanism. There is an implicit subsidy, which reduces the perceived marginal cost of provision of the regional public good by ½. Consequently, each region oversupplies the regional public good. Intuitively, each regional government anticipates that the center will redistribute consumption of the private good, equating individual marginal utilities of income, and thus has an incentive to overspend resources in consumption of the regional public good.

3.2.2 Selective Centralized Earmarking

The distortion created by the income transfer mechanism can be eliminated if the center introduces earmarking transfers to equalize fiscal capacities. Suppose now that the center earmarks the provision of regional public goods. As before, y_i denotes the amount of regional public good that region i provides. Let $e_i + s_i$ denote the total tax revenue that region i has available to provide the regional public good, where e_i represents the portion of tax revenue that is collected in the region and s_i is the amount of a (positive or negative) fiscal transfer promoted by the center. Since the fiscal transfer is earmarked, the regional government must balance its budget with respect to provision of the regional public good: $y_i = e_i + s_i$, $i = 1, 2$. This implies that region i's budget constraint becomes

$$nx_i + g_i + e_i + s_i = nw_i + \tau_i, \quad i = 1, 2. \tag{14}$$

Assume that the fiscal transfers are redistributive, so that

$$s_1 + s_2 = 0. \tag{15}$$

In this setting, regional government i chooses $\{e_i, g_i\}$ in the first stage. Having observed $\{g_1, g_2, e_1, e_2\}$, the center chooses $\{s_1, s_2, \tau_1, \tau_2\}$ to maximize social welfare (1) subject to constraints (9), (14), (15) and $y_i = e_i + s_i$, $i = 1, 2$. The conditions that maximize social welfare are the constraints, equation (10) and the following:

$$r'(y_1) = r'(y_2). \tag{16}$$

As before, equation (10) implies that $x_1 = x_2 = x$. Equation (16) informs us that the fiscal transfers equalize the marginal social utilities of consumption of regional public goods. Since $r'' < 0$, this equation implies that $y_1 = y_2 = y$. Let $x(e_1, e_2, g_1, g_2)$ and $s^i(e_1, e_2)$, $i = 1, 2$, denote the center's best response functions. These functions satisfy the following system of equations:

$$x(e_1, e_2, g_1, g_2) = \frac{W - E - G}{N}, \tag{17}$$

$$e_1 + s^1(e_1, e_2) = e_2 + s^2(e_1, e_2), \tag{18}$$

$$s^1(e_1, e_2) + s^2(e_1, e_2) = 0, \tag{19}$$

where $E = e_1 + e_2$ is the national expenditure incurred in the provision of regional public goods. Equations (18) and (19) imply that

$$\frac{\partial s_1}{\partial e_1} = \frac{\partial s_2}{\partial e_2} = -\frac{1}{2}. \tag{20}$$

In the first stage, regional government i chooses non-negative $\{e_i, g_i\}$ to maximize $n\left[b(x(e_1, e_2, g_1, g_2)) + r\left(e_i + s^i(e_1, e_2)\right) + v(f(g_1, g_2))\right]$, taking the choices of the other regional government as given. Assuming interior solutions, the first order conditions are (for $i = 1, 2$):

$$b'(x_i)\frac{\partial x}{\partial e_i} + r'(y_i)\left(1 + \frac{\partial s^i}{\partial e_i}\right) = 0 \;\Rightarrow\; \frac{N}{2}\frac{r'(y_i)}{b'(x_i)} = 1 \;\Rightarrow\; \frac{nr'(y_i)}{b'(x_i)} = 1, \tag{21}$$

$$b'(x_i)\frac{\partial x}{\partial g_i} + f_i v'(Q) = 0 \;\Rightarrow\; \frac{Nf_i v'(Q)}{b'(x_i)} = 1. \tag{22}$$

As revealed by equations (20) and (21), the fiscal transfer mechanism neutralizes the distortion created by the income transfer mechanism on the provision of regional public goods. As before, the regions face correct incentives for the contributions to the federal public good, as shown by equation (22).

3.2.3 Selective Decentralized Leadership

In order to formally demonstrate that the distortion created by the income transfer mechanism arises only if the regional governments are able to commit to their contributions to regional public goods, suppose now that regional governments are unable to commit to provision of regional public goods. In the first stage of the game, they choose their contributions to the federal public good. In the second stage of the game, the regional governments choose the amounts of regional public goods and the center chooses the amounts of private goods.

Consider the second stage. Given equations (8), one can say that regional government i chooses non-negative y_i to maximize $n\left[b\left(w_i - \frac{g_i+y_i-\tau_i}{n}\right) + r\left(y_i\right) + v(Q)\right]$, taking the choices of the other governments as given. In addition, the central government chooses $\{\tau_1, \tau_2\}$ to maximize social welfare, $n\Big[b\Big(w_1 - \frac{g_1+y_1-\tau_1}{n}\Big) + b\Big(w_2 - \frac{g_2+y_2-\tau_2}{n}\Big) + r(y_1) + r(y_2) + 2v(Q)\Big]$, subject to the income redistribution constraint (9), taking the choices of the other governments as given. Assuming interiors solution, the first order conditions yield equations (6), (9) and (10). Combining equations (8) and (10) yields the national resource constraint (2).

Let $y^i(g_1, g_2)$ and $\tau^i(g_1, g_2)$ denote the best response functions for the regional and central governments, respectively. Let

$$x^i\left(g_1, g_2\right) = w_i + \frac{\tau^i\left(g_1, g_2\right) - g_i - y^i\left(g_1, g_2\right)}{n}. \tag{23}$$

Equation (10) implies that $x^1\left(g_1, g_2\right) = x^2\left(g_1, g_2\right)$. Hence, this result and equations (23) imply

$$w_1 + \frac{\tau^1\left(g_1, g_2\right) - g_1 - y^1\left(g_1, g_2\right)}{n} = w_2 + \frac{\tau^2\left(g_1, g_2\right) - g_2 - y^2\left(g_1, g_2\right)}{n}. \tag{24}$$

Equation (9) yields

$$\tau^1\left(g_1, g_2\right) + \tau^2\left(g_1, g_2\right) = 0. \tag{25}$$

Equations (24) and (25) yield

$$\tau^1\left(g_1, g_2\right) = \frac{n\left(w_2 - w_1\right) + \left(g_1 + y^1\left(g_1, g_2\right) - g_2 - y^2\left(g_1, g_2\right)\right)}{2} = -\tau^2\left(g_1, g_2\right). \tag{26}$$

Letting $x^i\left(g_1, g_2\right) = x\left(g_1, g_2\right)$ and combining equations (23) and (26) implies

$$x\left(g_1, g_2\right) = \frac{W - G - Y\left(g_1, g_2\right)}{N} = \frac{W - G - 2y\left(g_1, g_2\right)}{N}. \tag{27}$$

Equation (27) uses the fact that $Y(g_1, g_2) = y^1(g_1, g_2) + y^2(g_1, g_2) = 2y(g_1, g_2)$, where the last equality follows from the combination of equations (6) and (10), which yields $y^1(g_1, g_2) = y^2(g_1, g_2) = y(g_1, g_2)$.

Consider now the first stage. Regional government i chooses non-negative g_i to maximize $n[b(x(g_1, g_2)) + r(y(g_1, g_2)) + v(f(g_1, g_2))]$, taking the choice of the other regional government as given. Assuming interior solutions, the first order conditions in the first stage are (for $i = 1, 2$):

$$b'(x)\frac{\partial x}{\partial g_i} + r'(y)\frac{\partial y}{\partial g_i} + f_i v'(Q) = 0 \;\Rightarrow\; \frac{\partial y}{\partial g_i}\left(r'(y) - \frac{2b'(x)}{N}\right) - \frac{b'(x)}{N} + f_i v'(Q) = 0$$
$$\Rightarrow\; \frac{2b'(x)}{N}\frac{\partial y}{\partial g_i}\left(\frac{nr'(y)}{b'(x)} - 1\right) + \frac{b'(x)}{N}\left(\frac{Nf_i v'(Q)}{b'(x)} - 1\right) = 0 \;\Rightarrow\; \frac{Nf_i v'(Q)}{b'(x)} = 1.$$

In sum, the subgame perfect equilibrium for the selective decentralized leadership game is socially optimal.

4 Economy with Mobile Residents

Having considered the special case in which residents are immobile and the regions are equally populated, suppose now that individuals are free to choose their region of residence. Every individual is endowed with one unit of labor, which he/she supplies to competitive firms that produce the numeraire good. Assume that labor is the only variable input. There are $J_i \geq 2$ firms in region i. Let j index the firms, with $j = 1, \ldots, J_i$ in region i. The firms in both regions use a standard, constant-returns-to-scale technology. Let $\Phi(.)$ denote the concave production function that represents this technology. Assume that $\Phi(.)$ increases in all arguments at decreasing rates and assume that all inputs are essential (i.e., $\Phi(.)$ satisfies the standard Inada conditions).[4] Let $\phi^i(l_{ji}) \equiv \Phi(l_{ji}; \mathbf{z}_{ji})$ be the reduced form for the production function utilized by the representative firm in region i, where l_{ji} is the amount of labor that the firm hires and $\mathbf{z}_{ji}$ is the vector of fixed inputs. Assume that all firms in region i use the same quantities of fixed inputs; that is, $\mathbf{z}_{ji} = \mathbf{z}_i$, for all $j = 1, \ldots, J_i$. In addition, assume that $\mathbf{z}_1 >> \mathbf{z}_2$ and $J_1 \geq J_2$. Region 1 is more abundant in the fixed resources than region 2. This is the sole source of asymmetry in the model.

All firms operate in a common labor market. Let ω denote the market wage. The representative firm in region i chooses $l_{ji} \geq 0$ to maximize $\phi^i(l_{ji}) - \omega l_{ji}$, taking the choices of all other firms as given. The first order conditions yield $\phi^i_{l_{ij}} = \omega, i = 1, 2$, where $\phi^i_{l_{ij}} \equiv d\phi^i / dl_{ij}$. The amount of labor hired by the representative firm satisfies the equalization of the marginal product of labor to the marginal cost of labor. Let $l^i_j(\omega)$ denote the labor demand function for the representative firm in region i. Note

[4]The Inada conditions guarantee that both regions are populated in equilibrium.

that $l_j^i(\omega) = l^i(\omega)$, for all $j = 1, \ldots, J_i$. In words, all firms in region i hire the same amount of labor.

The labor market clears if and only if

$$L^1(\omega) + L^2(\omega) = N, \tag{28}$$

where $L^i(\omega) \equiv J_i l^i(\omega)$, $i = 1, 2$. The market-clearing condition (28) can be used to define the market wage as an implicit function of the labor market characteristics, $\omega = \omega(J_1, J_2, N)$.

The representative consumer in region i earns an amount w_i of market income from supplying labor in the market and from being a shareholder in all firms located in the region.[5] Then, we can write the per capita market income function in region i as follows:

$$w^i(n_i, \omega) = \omega + J_i \pi^i(\omega)/n_i, \quad i = 1, 2, \tag{29}$$

where $\pi^i(\omega) \equiv f\left(l^i(\omega)\right) - \omega l^i(\omega)$ is the profit obtained by the representative firm in region i. In addition to market income, the representative resident of region i also receives (if positive) or pays (if negative) a transfer of τ_i/n_i units of income from the central government. This consumer spends $x_i + t_i$ units of income to pay for his/her private and public consumption levels, where t_i is the amount of tax that he/she pays to the regional government. Hence, the consumer's budget constraint yields:

$$x^i(n_i, t_i, \tau_i) = w^i(n_i, \omega) - t_i + \tau_i/n_i, \quad i = 1, 2. \tag{30}$$

Regional government i must balance the regional public budget:

$$t^i(g_i, n_i, y_i) = (g_i + y_i)/n_i, \quad i = 1, 2, \tag{31}$$

where $t^i(g_i, n_i, y_i)$ is region i's per capita tax function. The central government's income transfers are redistributive. Hence, constraint (9) holds.

Due to idiosyncratic regional benefits (e.g., family ties, culture, language, etc.), consumers are attached to regions. Let $n \in [0, N]$ denote a consumer in the economy. This individual gets an attachment benefit equal to $a(N - n)$ if he/she resides in region 1 and an attachment benefit equal to an if he/she resides in region 2, where $a > 0$ is the attachment intensity. The total utility individual n derives from residing in region 1 is $u(x_1, y_1, Q) + a(N - n)$, while the total utility this individual derives from residing in region 2 is $u(x_2, y_2, Q) + an$. In the migration equilibrium, there

[5]For simplicity, I omit rental income sources that residents obtain from supplying fixed inputs (say, land and capital) in the market. If each resident of region i is endowed with equal supplies of region i's fixed resources, equation (28) would remain the same: the rental incomes would exactly cancel out with the amounts spent by the firms to hire such resources. Thus, the analysis in the text is consistent with the assumption that regional residents are equally endowed with all regional resources.

is an individual, n_1, who is indifferent between residing in region 1 and residing in region 2:

$$\begin{aligned} &u(x_1, y_1, Q) + a(N - n_1) = u(x_2, y_2, Q) + an_1 \\ &\Rightarrow\ b(x_1) + r(y_1) + a(N - n_1) = b(x_2) + r(y_2) + an_1. \end{aligned} \tag{32}$$

4.1 Social Optimum

Let $U^1 = \int_0^{n_1} [b(x_1) + r(y_1) + v(Q) + a(N-n)]dn$ and $U^2 = \int_{n_1}^{N} \left[b(x_2) + r(y_2) + v(Q) + an\right]dn$ be the welfare levels enjoyed by regions 1 and 2, respectively. We assume that the social planner is utilitarian, $U = U^1 + U^2$. Hence,

$$\begin{aligned} U =& n_1\left[b(x_1) + r(y_1) + v(Q) + a\left(N - \frac{n_1}{2}\right)\right] \\ &+ n_2\left[b(x_2) + r(y_2) + v(Q) + a\left(N - \frac{n_2}{2}\right)\right]. \end{aligned} \tag{33}$$

Assume that consumers/migrants do not observe the social planner's policy choices prior to making their migration decisions and the social planner does not observe the outcome of the migration decisions (i.e., the population distribution) prior to making its policy choices.[6] These assumptions imply that consumers/migrants take the planner's choices as given while the planner takes the population distribution as given. The migration equilibrium is determined by equation (33). Taking $\{n_1, n_2\}$ as given, the planner chooses $\{g_1, g_2, y_1, y_2, \tau_1, \tau_2\}$ to maximize

$$\begin{aligned} &n_1\left[b\left(w^1(n_1, \omega) + \frac{\tau_1 - g_1 - y_1}{n_1}\right) + r(y_1)\right] \\ &+ n_2\left[b\left(w^2(n_2, \omega) + \frac{\tau_2 - g_2 - y_2}{n_2}\right) + r(y_2)\right] + Nv(Q), \end{aligned} \tag{34}$$

subject to constraint (9), where the objective function (34) neglects the attachment benefits present in the social welfare function (33) because the planner takes the population distribution as given. In writing (34), equations (29) and (30) are taken into account. Remember that the per capita market income functions are given by equations (28). Straightforward calculations yield conditions (7), (10) and the

[6] Silva and Lucas (2016) show that the subgame perfect equilibrium for the decentralized leadership game is the same whether government authorities take migration responses into account or consider the population distribution as given. To simplify exposition, the government authorities take the population distribution as given in the current setting.

following:

$$\frac{n_i r'(y_i)}{b'(x_i)} = 1, \quad i = 1, 2, \tag{35}$$

$$n_1 x_1 + n_2 x_2 + Y + G = W, \tag{36}$$

where, in the national resource constraint (36), $W = n_1 w^1(n_1, \omega) + n_2 w^2(n_2, \omega)$. Since the objective function is strictly concave and the constraint (9) is linear, the solution to the planner's problem satisfies the sufficient second order conditions and it is unique.

Combining equations with the migration equilibrium equation (32) and $N = n_1 + n_2$ yields the socially optimal allocation subject to free mobility of residents. Equation (10) implies that $x_1 = x_2 = x$. Combining equations (10) and (35), one obtains

$$n_1 r'(y_1) = n_2 r'(y_2). \tag{37}$$

The fact that $x_1 = x_2 = x$ implies that the migration equilibrium equation (32) simplifies to

$$r(y_1) + a n_2 = r(y_2) + a n_1. \tag{38}$$

For arbitrary a and N values, equations (37), (38) and $N = n_1 + n_2$ hold simultaneously if and only if $n_1 = n_2 = n = N/2$ and $y_1 = y_2 = y$.

4.2 Unlimited Decentralized Leadership with Selective Earmarking

Suppose that the regional governments are able to commit to their contributions to regional and federal public goods. The center, however, is able to make income and fiscal transfers. The fiscal transfer that a region receives or pays is earmarked. As in Sect. 3.2.2, let $y_i = e_i + s_i$, $i = 1, 2$. Assume also that $s_1 + s_2 = 0$. The budget constraint for the representative resident of region i yields

$$x^i(e_i, g_i, n_i, s_i, \tau_i) = w^i(n_i, \omega) + \frac{\tau_i - g_i - e_i - s_i}{n_i}, \quad i = 1, 2. \tag{39}$$

Having observed $\{e_1, e_2, g_1, g_2\}$, the center chooses $\{s_1, \tau_1\}$ to maximize

$$n_1 \left[b \left(x^1 \left(e_1, g_1, n_1, s_1, \tau_1 \right) \right) + r \left(e_1 + s_1 \right) \right]$$
$$+ n_2 \left[b \left(x^2 \left(e_2, g_2, n_2, -s_1, -\tau_1 \right) \right) + r \left(e_2 - s_1 \right) \right],$$

in the second stage, taking $\{n_1, n_2\}$ as given. The objective function takes $s_2 = -s_1$ and $\tau_2 = -\tau_1$ into account and ignores the national benefit from consumption of the federal public good, since Q has already been determined in the first stage. The first order conditions yield equations (10) and (37). Let $s^i(e_1, e_2, g_1, g_2)$ and $\tau^i(e_1, e_2, g_1, g_2)$ denote the center's best response functions, $i = 1, 2$. Equation (10) yields

$$w^1 \left(n_1, \omega \right) + \frac{\tau^1(.) - g_1 - e_1 - s^1(.)}{n_1} = w^2 \left(n_2, \omega \right) + \frac{\tau^2(.) - g_2 - e_2 - s^2(.)}{n_2}. \tag{40}$$

Plugging $\tau^2(.) = -\tau^1(.)$ into equation (40) and solving the resulting expression yields

$$\tau^1(.) = \frac{n_1 n_2 \left[w^2 \left(n_2, \omega \right) - w^1 \left(n_1, \omega \right) \right] + n_2 \left[e_1 + g_1 + s^1(.) \right] - n_1 \left[e_2 + g_2 + s^2(.) \right]}{N}$$
$$= -\tau^2(.). \tag{41}$$

Combining equations (40) and (41) one obtains

$$\widehat{x} \left(e_1, e_2, g_1, g_2 \right) = \frac{W - E - G}{N}. \tag{42}$$

Equation (37) yields

$$n_1 r' \left(e_1 + s^1(.) \right) = n_2 r' \left(e_2 + s^2(.) \right). \tag{43}$$

Plugging $s^2(.) = -s^1(.)$ into equation (43) and differentiating with respect to each policy variable controlled by the regional governments leads to ($i = 1, 2$):

$$\frac{\partial s^i}{\partial e_i} = -\frac{n_i r'' \left(y_i \right)}{n_1 r'' \left(y_1 \right) + n_2 r'' \left(y_2 \right)} < 0, \tag{44}$$

$$\frac{\partial s^i}{\partial e_{-i}} = \frac{n_{-i} r'' \left(y_{-i} \right)}{n_1 r'' \left(y_1 \right) + n_2 r'' \left(y_2 \right)} > 0, \quad -i = 2 \quad \text{if} \quad i = 1 \quad \text{and vice versa}, \tag{45}$$

$$\frac{\partial s^i}{\partial g_i} = \frac{\partial s^i}{\partial g_{-i}} = 0, \quad -i = 2 \quad \text{if} \quad i = 1 \quad \text{and vice versa}. \tag{46}$$

In the first stage, regional government i chooses non-negative $\{e_i, g_i\}$ to maximize

$$n_i\left[b\left(\widehat{x}\left(e_1, e_2, g_1, g_2\right)\right) + r\left(e_i + s^i(.)\right) + v\left(f\left(g_1, g_2\right)\right)\right]$$

taking $\{n_1, n_2\}$ and the choices of the other regional government as given. Assuming interior solutions, the first order conditions yield equations (7) and the following (for $i = 1, 2$)

$$\frac{Nr'\left(y_i\right)}{b'(x)}\left(\frac{n_{-i}r''\left(y_{-i}\right)}{n_1 r''\left(y_1\right) + n_2 r''\left(y_2\right)}\right) = 1. \tag{47}$$

Since the equilibrium also satisfies equation (38) and $N = n_1 + n_2$, one obtains $n_1 = n_2 = n = N/2$ and $y_1 = y_2 = y$. Combining these results with equations (47) yields conditions (7).

5 Conclusion

In most nations, regional governments provide multiple public goods. Some of these goods yield regional consumption benefits while some others benefit residents and non-residents. In federal regimes where regional governments have some policy autonomy, the provision of regional and federal public goods may be efficient. This paper demonstrates that there are common circumstances under which regional and central governments interact in efficient manners in federations. In particular, regional and central governments may implement socially optimal policies if regional governments commit to provision of regional and federal public goods prior to the center's decisions concerning interregional income and earmarked transfers. The efficiency of decentralized leadership is robust to imperfect residential mobility if there is a common labor market in the federation.

Caplan et al. (2000) applied the rotten kid theorem obtained by Cornes and Silva (1999) to a federal regime context due to the similarities there exist between families with rotten kids and loving and benevolent parents and federations with self-interested regional governments and benevolent central governments. Caplan et al. (2000) was also motivated by the fact that individuals seem to be attached to regions and hence not perfectly mobile in many federations. In the current paper, the positive message of Caplan et al. (2000) is shown to hold to more general situations, which include multiple types of public goods, but under conditions that they did not anticipate. Without residential mobility and unlimited regional commitments, the efficiency result goes through only if the center possesses instruments to implement interregional income and earmarked transfers. With residential mobility subject to attachment to regions and unlimited regional commitments, the efficiency result goes through only if the center is endowed with the previously mentioned instruments and there is a common labor market in the federation.

Since the application to federal regimens has produced new insights, a natural question is to ask to what extent such insights are applicable to family contexts. Rotten kids consume a large and diverse basket of goods. Some of such goods are private goods which yield private benefits, others are private goods that produce (positive or negative) externalities while others still are family (i.e., public) goods. The insights from this paper generate the following behavioral hypotheses: rotten kids overconsume private goods other than the numeraire and provide efficient contributions to family goods in the presence of family income redistribution (e.g., inter-vivo transfers and bequests). As data about family expenditures become more abundant and available, such behavioral hypotheses should be tested. Another interesting avenue for future research in family contexts is to study interfamily interactions and the provision of multiples public goods, some of which are commonly shared [see, e.g., Cornes et al. (2012)]. In family networks, there may be special arrangements of income and earmarked transfers that motivate family members to be well behaved.

References

Becker, G. S. (1981). *A treatise on the family*. Cambridge, MA: Harvard University Press.

Bergstrom, T. (1989). A fresh look at the rotten kid theorem—And other household mysteries. *Journal of Political Economy, 97*, 1138–1159.

Caplan, A., Cornes, R. C., & Silva, E. C. D. (2000). Pure public goods and income redistribution in a federation with decentralized leadership and imperfect labor mobility. *Journal of Public Economics, 77*, 265–284.

Cornes, R. C. (1993). Dyke maintenance and other stories: Some neglected types of public goods. *Quarterly Journal of Economics, 108*, 259–271.

Cornes, R., & Hartley, R. (2007). Weak links, good shots and other public good games: Building on BBV. *Journal of Public Economics, 91*, 1684–1907.

Cornes, R. C., & Itaya, J.-I. (2010). On the private provision of two or more public goods. *Journal of Public Economic Theory, 12*, 363–385.

Cornes, R. C., Itaya, J.-I., & Tanaka, A. (2012). Private provision of public goods between families. *Journal of Population Economics, 25*, 1451–1480.

Cornes, R. C., & Silva, E. C. D. (1999). Rotten kids, purity and perfection. *Journal of Political Economy, 107*, 1034–1040.

Silva, E. C. D. (2014). Selective decentralized leadership. *Journal of Urban Economics, 83*, 1–5.

Silva, E. C. D. (2015). Efficient earmarking under decentralized fiscal commitments. *International Tax and Public Finance, 22*, 683–701.

Silva, E. C. D. (2016). Tax competition and federal equalization schemes with decentralized leadership. *International Tax and Public Finance.* doi:10.1007/s10797-016-9402-7

Silva, E. C. D., & Lucas, V. M. (2016). Common labor market, attachment and spillovers in a large metropolis,”. *International Tax and Public Finance, 23*, 693–715.

Debt Neutrality Without Altruism: Voluntary Contributions to Public Goods as 'Operative Linkages'

Christian Haslbeck and Wolfgang Peters

1 Introduction

Two central neutrality results have considerably influenced modern economic policy analysis. First, there is the *Ricardian Equivalence Theorem*. This theorem states that taxes and public debt are equivalent instruments of financing public expenditures. Since rational individuals fully anticipate that the present value of the future tax liabilities caused by the debt services is always equal to the current tax cut, their wealth is not affected by public debt. Therefore, they will choose identical consumption patterns under each regime.[1] This argument was criticized because it breaks down if the term of public debt exceeds individual lifetimes. Barro (1974) revived the Ricardian Equivalence Theorem in his path-breaking article. He argued that individuals, who are altruistic towards their direct descendants, will take into account the future tax liabilities of their children. For that reason parents will increase their bequests just by the amount necessary to finance future tax increases and bequest adjustments of their heirs. In this way Ricardian Equivalence is restored.

We have benefitted from valuable comments and criticism of an anonymous referee and by Dirk Rübbelke.

[1]For a discussion of the conditions under which Ricardian Equivalence is valid see, for example, Brennan and Buchanan (1980).

C. Haslbeck
Bavarian Ministry for Economic Affairs, Transportation and Technology, Prinzregentenstraße 28, 80538 Munich, Germany
e-mail: christian.haslbeck@stmwivt.bayern.de

W. Peters (✉)
Department of Economics, European University Viadrina, Große Scharrnstraße 59, 15230 Frankfurt (Oder), Germany
e-mail: Peters@europa-uni.de

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_10

The second neutrality theorem applies to the theory of voluntary provision of public goods. According to this theorem, which was recognized first by Becker (1974) and derived more formally by Warr (1983),[2] a lump-sum redistribution of incomes between contributors to the public good does not affect the allocation of resources in Nash equilibrium.[3] Warr (1983) showed that taxed (subsidized) individuals reduce (increase) their contributions to the public good just by the tax (subsidy). Thus, total public good supply, private consumption levels, and agents' welfare positions are unaffected by the redistributive policy.

At first sight these two neutrality theorems seem to have little in common. However, a more careful inspection reveals that they are related closer than it might appear. In particular, both theorems are concerned with the inefficacy of income redistributions. While Warr (1983) considers income redistributions between contributors to a public good, who live at the same time, the Ricardian Equivalence Theorem applies to income redistributions between generations living in different periods. The purpose of this paper is to demonstrate that in a model, in which a public good is financed through voluntary contributions by overlapping generations, Warr's neutrality result is also valid for intergenerational income redistributions. As a consequence, Ricardian Equivalence must hold as well.

The idea that Warr neutrality and Ricardian Equivalence have a common root is not new. Andreoni (1989) argued that, if individuals care about the consumption of their heir, this consumption can be interpreted as a public good. As a consequence, he viewed the problems of public good provision and intergenerational transfers as formally equivalent.[4] However, his model is essentially an application of the static standard model of public good provision to the case of transfers between two generations which cannot be adapted to a context with more than two succeeding generations.

Unlike the framework of Andreoni (1989), the model presented in this paper is genuinely dynamic. As a consequence, the linkages between succeeding generations become more complex. In addition and contrary to Andreoni (1989) and the literature on Ricardian Equivalence in general, there is no altruism embodied in our model. Nevertheless it can be shown that the voluntary provision of a public good provides exactly that kind of 'operative linkage' between generations which, according to Bernheim and Bagwell (1988), is suited to bring about Ricardian Equivalence. This is the central message of this paper.

[2]For a simpler and more elegant proof see Cornes and Sandler (1984).

[3]This neutrality result holds even for special cases of distortionary tax-subsidy schemes, see Bernheim (1986), Andreoni (1988) or Buchholz et al. (2006). Examples of non-neutral tax-subsidy schemes are provided by Boadway et al. (1989), Nett and Peters (1993), and Andreoni and Bergstrom (1996).

[4]See Rebelein (2002, 2006) for a similar line of reasoning. He presents a two-period, two-consumer model with one-sided altruism. While a child tries to manipulate the size of its parent's bequest strategically, the parent seeks to minimize this impact. He shows that even if parents as well as their children behave strategically, Ricardian Equivalence remains active.

The paper is organized as follows: In Sect. 2 we introduce the two-period overlapping-generations model in which young and old generations contribute to a public good. In Sect. 3 the game which represents the conflict between young and old generations is described. We work out the strategic effects of generations' present savings on their future private and public consumption levels. These effects play a key role for the understanding of the reason why linkages between generations are operative. The section finishes with the characterization of interior solutions to the development of public good supply and to the intergenerational cost sharing. In Sect. 4 we consider the intergenerational income redistribution resulting from a simple public debt policy. As our central result we prove that this policy is neutral with respect to the intertemporal allocation of resources. In Sect. 5 we discuss this result with reference to the existing literature and provide some concluding remarks.

2 The Model

In this section we develop the basic framework for the analysis of the intertemporal provision of a public good. For simplicity we consider an overlapping-generations model with only two generations, the 'young' and the 'old'. It is assumed that each generation's life cycle can be divided into two periods. In addition, we suppose that population is stationary and that all individuals are identical. Hence, each generation's preferences can be described by the utility function of a representative consumer. Utility of the generation born at time t is given by

$$u_t = u(c_t, z_{t+1}, Q_t, Q_{t+1}), \tag{1}$$

where c_t and z_{t+1} denote present and future private consumption, while Q_t and Q_{t+1} stand for present and future quantities of the public good.

Public good supply in each period amounts to the voluntary contributions of the currently living generations,

$$Q_t = p_t + q_t, \tag{2}$$

where p_t and q_t are the contributions of the young and the old generation, respectively.

Each young generation is endowed with a fixed, time independent labor income w, which can be spent on present consumption, c_t, used as a contribution to the public good, p_t, or as savings, s_t. Assuming that the price of the public good is constant and normalized to unity, the budget constraint for the first life phase is given by

$$w = c_t + p_t + s_t. \tag{3}$$

Savings can be invested at a constant market interest rate, r. When retired, a generation has no other sources apart from its accumulated wealth. Wealth is split between consumption, z_{t+1}, and a contribution to the public good, q_{t+1}. In particular, there are no bequests. Thus, the budget constraint in the second life phase is given by

$$(1 + r)s_t = z_{t+1} + q_{t+1}. \tag{4}$$

3 A Game of Intertemporal Public Good Provision

In this section we explain in detail the conflict between young and old generations, which drives the development of public good supply and intergenerational cost-sharing. We show that this development is essentially determined by the strategic effects of young generations' savings on their future private and public consumption levels.

As both generations (old and young) share the burden of providing the public good, at least the young generation can try to shift this burden towards the old one through saving a little bid more. In that case a lower current disposable income yields in less contributions while young. However, these increased savings enlarge disposable income in the next period. Thus, a young generation has to search for a balance in improving today's against future opportunities of burden shifting. The correct anticipation of these strategic effects by the young generations will be of central importance for the understanding of the intergenerational linkage which is decisive for our neutrality result.

3.1 The Structure of the Intergenerational Conflict

The structure of the non-cooperative game in which the contributions of the young and the old generations to the public good are determined is shown in Fig. 1. Each row of this figure displays the decisions of a single generation made during its life cycle. Each column shows the decisions of the two different generations living in the same period.

First note that each generation plays twice. Refer to the generation born at time t as to 'generation t' and consider the game played in the initial period 1. In this period the young generation 1 plays against the currently old generation 0. In period 2, generation 1 has grown old itself and plays against the now young generation 2. Thus each generation is involved in a two-stage game with different opponents at the first and the second stage. Note that this is not a typical repeated game. In a typical repeated game an unchanged set of players plays the same stage game for several times. In our model, by contrast, two generations meet only once and never

Fig. 1 The structure of intergenerational conflict

	1	2	period $\longrightarrow$	$T-1$	T
0	q_1				
1	(p_1, s_1)	q_2			
2		(p_2, s_2)	$\dots$		
$\downarrow$					
$T-2$			$\dots$	q_{T-1}	
$T-1$				(p_{T-1}, s_{T-1})	q_T
T					(p_T, s_T)
generation					

again. In the next period the old generation is replaced by a new young generation and both play a possibly different game. We assume that the entire game ends at a predetermined point in time, T. Think, for example, of a public good provided by the members of a community of interests which has committed itself in its statutes to dissolve in T.

Now suppose that in each period the decisions of the old and the young generation are made simultaneously and independently. Hence, we consider Nash equilibria in each period. Since generations play non-cooperatively against each other, they have to solve a conflict in financing the public good. Therefore, unlike in Barro (1974), they can not be treated as one 'dynastic' family represented by a single agent who lives from period 1 to T and decides about the uses of total intertemporal resources. Thus, the outcome of the entire game is certainly inefficient, as in almost any non-cooperative game on voluntary provision of a public good.[5]

Note in addition that, in our model, it is not likely that inefficiency can be avoided by enforcing cooperation between generations through strategies of punishment or reputation. Clearly cooperation fails, due to the single confrontation, a defecting generation can never be punished directly by the deceived generation. Eventually one might imagine a kind of social contract threatening to punish 'unfriendly' behavior of a generation in its first life phase by a corresponding behavior of the next generation.[6] However, such a punishment could be inflicted only on young generations, while old generations could always defect unpunished. Therefore, punishment cannot bring about full cooperation. The same is true for strategies of reputation. As everyone knows, an old generation will always defect from cooperation, no matter how it acted when young. Therefore, no generation has an incentive in the first life phase to build up the reputation of being cooperative. Thus, each generation will defect in either period. Thus, non-cooperative behavior prevails.

Given these general remarks consider now the generations' decision problems in period t. The old generation chooses its contribution q_t optimally, given its savings

[5]Cf. Cornes and Sandler (1984), Bergstrom et al. (1986) or Bernheim (1986).

[6]For the idea of social contracts to enforce cooperation between generations see Kotlikoff et al. (1988).

from the preceding period, s_{t-1}, and the contribution of the young generation, p_t. From this choice we obtain the old generation's *reaction function*, $q_t(p_t, s_{t-1})$. Private consumption z_t follows as a residual. The young generation chooses p_t and s_t, given its wage, w, and the old generation's contribution, q_t. This choice yields *two* reaction functions $p_t(q_t; w)$ and $s_t(q_t; w)$ of the young generation. Again, private consumption c_t follows as a residual. Using the reaction functions we can determine the Nash equilibrium. Note that the equilibrium values depend all on the savings s_{t-1}. Just as the Nash equilibrium of period t depends on s_{t-1}, the equilibrium of any period depends on the savings of the preceding period. Thus, to make sure that its decision will be time consistent, the generation which is young in t has to anticipate the strategic effects of savings on the equilibrium in period $t + 1$. These strategic effects are formally described by functions $z_{t+1}(s_t)$ and $Q_{t+1}(s_t)$ indicating how future private and public consumption levels vary with respect to changes in present savings.

Note that, to work out $z_{t+1}(s_t)$ and $Q_{t+1}(s_t)$, the young generation must anticipate correctly all future Nash equilibria. This is so, because public good supply in $t + 1$ is partly determined by p_{t+1}, the contribution of the generation which is young, then. However, by the same reasoning as above, to be able to predict p_{t+1} it is necessary to know how the equilibrium in $t + 2$ depends on s_{t+1}. To know this, one must know p_{t+2}, which in turn requires to know how the equilibrium in $t + 3$ depends on s_{t+2}, and so on. Given that the game is finite, $z_{t+1}(s_t)$ and $Q_{t+1}(s_t)$ can be obtained by backward induction. Since in the terminal period T the public good is supplied and consumed for the last time, the generation which is young in this period need not worry about the future strategic effects of its savings. Therefore, it is possible to calculate the Nash equilibrium of T as a function of past variables alone. Given this, we can determine all stage equilibria back to any arbitrary period t. This procedure yields the subgame-perfect solution of the entire game and results in equilibrium sequences $\{p_t; q_t; Q_t\}_{t=1,\dots,T}$ indicating how public good supply and intergenerational cost-sharing develop over time.

3.2 The Strategic Effects of Saving

Subsequently, we describe how the subgame-perfect equilibrium of the entire game can be determined. Given a general utility function (1), we could only make statements about the existence of subgame-perfect Nash equilibria. However, we want to carry out backward induction explicitly as explained above in order to obtain qualitative results concerning the intertemporal resource allocation. Only qualitative results will enable us to prove the ineffectiveness of intergenerational income redistributions. Hence we specify

$$u(c_t, z_{t+1}, Q_t, Q_{t+1}) = \ln c_t + \alpha \ln Q_t + \beta \left[\ln z_{t+1} + \alpha \ln Q_{t+1}\right]. \tag{5}$$

To solve the game for this utility function we have to proceed basically as outlined above. Recall that subgame-perfectness requires that the young generation takes into account the strategic effects of savings s_t on the Nash equilibrium in period $t+1$. The following proposition summarizes the results concerning these strategic effects:

Proposition 1 *For each generation* $t = 1, \dots, T-1$ *the strategic effects of present savings on future private and public consumption are described by*

$$\alpha z_{t+1} = Q_{t+1} = a\left[(1+r)s_t + \sum_{i=t+1}^{T} \frac{w}{(1+r)^{i-(t+1)}}\right], \tag{6}$$

where $a = \alpha/[(1+\alpha)(1+\beta)+1]$.

Proof The proof is given by complete induction. Using backward induction we begin with calculating the Nash equilibrium of the terminal period T. The old generation's decision problem in this period is

$$\max_{q_T} \ln\left[(1+r)s_{T-1} - q_T\right] + \alpha \ln\left[p_T + q_T\right]. \tag{7}$$

Since the project ends in T, the generation which is young in this period has no opportunity to consume the public good when old. Thus, its optimization problem reduces to

$$\max_{p_T, s_T} \ln\left[w - p_T - s_T\right] + \alpha \ln\left[p_T + q_T\right] + \beta \ln\left[s_T(1+r)\right]. \tag{8}$$

Solving these two problems, we obtain

$$\alpha z_T = Q_T = \frac{\alpha}{2+\alpha+\beta}\left[w + (1+r)s_{T-1}\right], \tag{9}$$

which indicates how old age private and public consumption in Nash equilibrium depend on the savings of the preceding period. Now consider the corresponding decision problems for the preceding period $T-1$. The old generation's problem is

$$\max_{q_{T-1}} \ln\left[(1+r)s_{T-2} - q_{T-1}\right] + \alpha \ln\left[p_{T-1} + q_{T-1}\right]. \tag{10}$$

To obtain a time consistent consumption plan, the young generation has to impose (9) on its decision problem. Given this, the young generation's problem can be written as

$$\max_{s_{T-1},p_{T-1}} \ln[w - p_{T-1} - s_{T-1}] + \alpha \ln[p_{T-1} + q_{T-1}] + \beta \ln\left[\frac{Q_T^{1+\alpha}}{\alpha}\right], \tag{11}$$

where $Q_T = \frac{\alpha}{2+\alpha+\beta}[w + (1+r)s_{T-1}]$. The solution of these two problems is straightforward and results in

$$\alpha z_T = Q_T = a\left[(1+r)s_{T-2} + w\left(1 + \frac{1}{1+r}\right)\right], \tag{12}$$

with $a = \alpha/[(1+\alpha)(1+\beta)+1]$. Obviously the condition $\alpha z_t = Q_t$ has to hold for any arbitrary period. This proves the first equality in Proposition 1. Furthermore, Eq. (12) suggests that the function which describes equilibrium public good supply at time t as a function of s_{t-1} is determined by a difference equation of the general structure

$$Q_t = a(1+r)s_{t-1} + k_t. \tag{13}$$

To prove this we show that if $Q_{t+1} = a(1+r)s_t + k_{t+1}$ holds, then (13) is valid. For that purpose we solve the problem

$$\max_{s_t,p_t} \ln[w - p_t - s_t] + \alpha \ln[p_t + q_t] + \beta \ln\left[\frac{Q_{t+1}^{1+\alpha}}{\alpha}\right], \tag{14}$$

where $Q_{t+1} = a(1+r)s_t + k_{t+1}$.

Simple calculus and some algebra lead us indeed to (13) with $k_t = \frac{k_{t+1}}{1+r} + aw$.This difference equation in combination with $k_{T-1} = aw[1 + 1/(1+r)]$, which is obvious from (12), enables us to calculate the entire time path for k_t:

$$k_t = aw\left(\frac{1+r}{r} - \frac{(1+r)^{t-T}}{r}\right) = a\sum_{i=t}^{T}\frac{w}{(1+r)^{i-t}}. \tag{15}$$

Using this equation in (13) completes the proof. ■

Proposition 1 states that, when old, generation t's private and public consumption in Nash equilibrium are equal to a constant share $a < 1$ of the economy's total wealth. Total wealth, which corresponds to the *present value of current and future resources available*, consists of two components: generation t's own wealth accumulated in the first life phase, and the present value of labor income of all future generations. The observation that generations calculate with the economy's total wealth as their 'disposable income' is central for our neutrality result.

3.3 Interior Solutions to Public Good Supply and Intergenerational Cost Sharing

Having characterized the strategic effects of young generations' savings, we can now proceed to analyze how public good supply develops over time and how the burden of financing the good is shared between young and old generations. Thereby we consider interior solutions exclusively, that is, we restrict the analysis on Nash equilibria in which both generations make strictly positive contributions in each period. This restriction is necessary for the following reason: From Bergstrom et al. (1986) we know that in a static public good model an interpersonal redistribution of incomes is non-neutral if the policy affects individuals who contribute nothing. Therefore, we can expect that the same will hold for the case of an intergenerational redistribution. Thus, to avoid that neutrality is precluded a priori, we have to concentrate on interior solutions in the entire time interval, in which the redistributive policy is active.

Using (5) and (6) together with the budget constraints (3) and (4) we can now formulate the decision problems which determine the Nash equilibrium in period t. Given s_{t-1} and p_t, the old generation chooses q_t maximizing

$$\ln\left[(1+r)s_{t-1}-q_t\right]+\alpha\ln(p_t+q_t). \tag{16}$$

Given q_t, the young generation chooses p_t and s_t maximizing

$$\ln(w-p_t-s_t)+\alpha\ln(p_t+q_t)+\beta\ln\left[\frac{Q_{t+1}^{1+\alpha}}{\alpha}\right]$$

$$\text{s.t.}Q_{t+1}=a\left[(1+r)s_t+\sum_{i=t+1}^{T}\frac{w}{(1+r)^{i-(t+1)}}\right]. \tag{17}$$

The solution of these two problems defines the Nash equilibrium in period t. For an interior equilibrium with non-negative contributions of both, the young and the old generation, we obtain the first order conditions

$$\frac{\alpha z_t}{Q_t}=1,\quad \frac{Q_{t+1}}{\beta(1+\alpha)c_t}=a(1+r),\quad \text{and}\ \frac{\alpha Q_{t+1}}{\beta(1+\alpha)Q_t}=a(1+r). \tag{18}$$

The first condition simply implies that for the old generation the marginal rate of substitution (MRS) between private and public consumption has to equal the corresponding relative price.[7] The interpretation of the other two conditions describing optimal behavior of the young generation is similar. Applying $\alpha z_{t+1} = Q_{t+1}$ to the utility function, the left-hand side of the second condition is the MRS between present private and future public consumption. The right-hand side is the corresponding relative price: If the young generation wants to increase present private consumption by one additional unit it has to reduce savings correspondingly, p_t held constant. This will reduce the economy's wealth in the next period by $1 + r$ units. According to (6), future public consumption in Nash equilibrium will be lower by a share a of this wealth reduction. Now consider the third condition of (18). The left-hand side denotes the MRS between present and future public consumption. Again, the right-hand side is the corresponding relative price: Increasing the contribution to present public consumption by one unit, the young generation's savings must decrease by the same amount, c_t held constant. This has the same effects on future wealth and future public consumption in Nash equilibrium as just described.

The interpretation of the share a becomes clearer if we reconsider that, according to (6), generations calculate with the economy's total wealth in period t as their disposable income. This income can be allocated on private and public consumption. Thus, we have

$$(1+r)s_{t-1} + \sum_{i=t}^{T} \frac{w}{(1+r)^{i-t}} = c_t + z_t + s_t + Q_t \tag{19}$$

as the budget constraint in period t. Substituting for the variables on the right-hand side from the optimality conditions (18) and differentiating with respect to Q_t yields just the budget share a spent on Q_t.[8]

Using the third condition of (18) we can now determine the time path of total public good supply. This condition is a difference equation which can be solved if an initial value for public good supply, Q_0, is given. We obtain

$$Q_t = Q_0 \lambda^t, \tag{20}$$

where $\lambda = [(1+r)(1+\alpha)\beta]/[(1+\alpha)(1+\beta)+1] > 0$.

Equation (20) implies that total public good supply increases (remains constant, decreases) over time if $\lambda > (=, <) 1$. Consider for example $\lambda > 1$. In this case the marginal cost of an additional unit of present in terms of future public consumption is 'relatively high'. Thus, according to the third optimality condition in (18), generations wish that public good supply should be higher in the second

[7] This is the typical characterization of individually rational behavior in the presence of a public good. Collectively rational behavior would require the adjustment of the sum of the young and the old generation's MRS to the relative price of the public good.

[8] This is known as a standard result from demand theory with logarithmic utility functions.

than in the first life phase. Therefore, public good supply must increase over time. The cases $\lambda = 1$ and $\lambda < 1$ can be interpreted in an analogous way.

The intergenerational cost sharing of financing the public good is obtained from the budget constraints (3) and (4), using the first order conditions (18) and inserting (20):

$$p_t = -\phi Q_0 \lambda^t + \sum_{i=t}^{T} \frac{w}{(1+r)^{i-t}},$$
$$q_t = (1+\phi) Q_0 \lambda^t - \sum_{i=t}^{T} \frac{w}{(1+r)^{i-t}}, \tag{21}$$

where $\phi = [1+\beta(1+\alpha)]/\alpha > 0$. Equations (18), (20) and (21) describe a complete interior solution for the equilibrium sequences $\{c_t, z_t, p_t, q_t, Q_t\}_{t=1,\dots,T}$. Savings s_t follow as a residual from one of the budget constraints.

4 Debt Financed Transfers

In this section we analyze the effects of governmental intervention on the provision level of the public good, on intergenerational cost sharing, and on the generations' welfare positions. Since the public good is provided in a non-cooperative game, its provision level will be inefficiently low. Therefore, the government might take into account some policy measure to increase public good supply. One policy the government might envisage is to grant in some period t one generation a once-and-for-all transfer, hoping that this will stimulate its willingness to contribute and thus cause at least a temporary improvement of the provision level. Alternatively, if the government does not rely on the citizens' private initiative, it can provide a part of the public good by itself paying a transfer in kind rather than a monetary one.

Concerning the question of financing the transfer there are two alternatives. First, a corresponding tax could be levied on the other generation living in t. However, assuming that all generations make positive contributions, we know from the standard model of voluntary provision of public goods that such a redistribution of incomes will be ineffective. This is the neutrality result of Warr (1983), which is directly applicable to our model if incomes are redistributed among individuals living in the same period. The second alternative is to finance the transfer via public debt. In this case a tax to pay the debt services has to be levied upon succeeding generations. One might expect that this kind of governmental intervention will lead to a temporary increase of public good supply in the transfer period, followed by future reductions of the taxed generations' contributions. Moreover, it seems plausible that, such a policy will favor the generation receiving the transfer, while

Table 1 A debt financed transfer with constant periodical debt services

Period	....	$t-2$	$t-1$	t	$t+1$	$t+2$	...	$T-1$	T
Tax	....	0	0	$-D$	τ	τ	...	τ	τ

the generations charged with the debt services will loose. Subsequently it will turn out that this intuition is not correct. We will show that even an intertemporal, and thus intergenerational income redistribution will neither affect public good supply, nor will it change the generations' welfare positions in any period.

4.1 Neutrality of Public Debt

Consider the simple debt policy illustrated in Table 1. Suppose that in period t the working generation is granted a transfer D,[9] which is financed by external debt bearing interest at the constant (world)market rate, r. Assume further that the debt has to be repaid until period T, and that in periods $t+1, \ldots, T$ a constant amount has to be brought up to finance debt redemption and interest payments.[10] For this purpose a lump-sum tax is levied upon each future working generation. This tax must equal the corresponding annuity,

$$\tau = \frac{rD}{1-(1+r)^{t-T}}. \tag{22}$$

Given this tax-transfer scheme, disposable income of generation t is now equal to $\tilde{w}_t = w + D$, while the generations living in the post-transfer periods earn $\tilde{w}_i = w - \tau$, $i = t+1, \ldots, T$. However, neither public nor private consumption is affected by these changes of disposable income. From (19) it is immediately clear that total public good supply is independent of the intertemporal income distribution. Moreover, the conditions in (18) imply that the private consumption levels c_t and z_t are proportional to total public good supply and thus remain unchanged as well. Consequently, since the budget constraints must be fulfilled in each period, the sole effects which can be caused by an intergenerational redistribution of incomes are adjustments of savings and changes of the intergenerational cost sharing. Subsequently it will be investigated in more detail how the debt policy is neutralized by generations' reactions.

[9] We assume that D is sufficiently small as to guarantee interior Nash equilibria in all periods.

[10] The assumption that the term of the credit ends in the same period as public good provision is not critical. Our results would be preserved for any interval $[t+1, S]$ with $S \leq T$.

4.2 How Public Debt is Neutralized

First note that in periods $1, \ldots t$ the present value of the economy's future income stream is not altered by the debt policy. Given the new intertemporal income profile, this present value amounts to

$$\sum_{i=t}^{T} \frac{\tilde{w}_i}{(1+r)^{i-t}} = \sum_{i=t}^{T} \frac{w}{(1+r)^{i-t}} + \left(D - \sum_{i=t+1}^{T} \frac{\tau}{(1+r)^{i-t}}\right). \tag{23}$$

Obviously the term in parentheses on the right-hand side of (23), which corresponds to the 'present value of the debt policy', is zero by the definition of the annuity τ in (22). This implies that, according to (21), generations' contributions are unaffected up to the transfer period t.

Beginning with the first post-transfer period, $t+1$, public good supply is the same as it would be in the absence of the policy. However, the cost sharing is different. It is straightforward to show that the contributions of the young and the old generations in the time interval $[t+1, T]$, are given by

$$\begin{aligned} \tilde{p}_k &= p_k - (1+r)D\frac{1-(1+r)^{k-T-1}}{1-(1+r)^{t-T}}, \quad k = t+1, \ldots T, \\ \tilde{q}_k &= q_k + (1+r)D\frac{1-(1+r)^{k-T-1}}{1-(1+r)^{t-T}}, \quad k = t+1, \ldots T, \end{aligned} \tag{24}$$

where p_k and q_k are the contributions in the absence of the policy as determined by (21).

Given (23), we can now describe the changes of generations' behavior which lead to the neutralization of public debt. These changes are summarized in Table 2.

Consider first what happens in period T, in which the tax τ is collected for the last time. The young generation T reacts to the tax with a reduction of its contribution to

Table 2 Neutralization of the debt policy

Period	'Young' generation	'Old' generation
t	$\Delta w_t = D$ $\Delta c_t = 0$ $\Delta p_t = 0$ $\Delta s_t = D$	$\Delta W_t = 0$ $\Delta z_t = 0$ $\Delta q_t = 0$
$k \in [t+1; T]$	$\Delta w_k = -\tau = -rD\frac{1}{1-(1+r)^{t-T}}$ $\Delta c_k = 0$ $\Delta p_k = -(1+r)D\frac{1-(1+r)^{k-T-1}}{1-(1+r)^{t-T}}$ $\Delta s_k = D\frac{1-(1+r)^{k-T}}{1-(1+r)^{t-T}}$	$\Delta W_k = (1+r)D\frac{1-(1+r)^{k-T-1}}{1-(1+r)^{t-T}}$ $\Delta z_k = 0$ $\Delta q_k = (1+r)D\frac{1-(1+r)^{k-T-1}}{1-(1+r)^{t-T}}$

the public good of equal amount, while consumption and savings are held constant. This can be seen in Table 2 if k is set equal to T. Anticipating this reaction, the old generation $T-1$ raises savings in its first life phase by an amount Δs_{T-1} such that the corresponding increase in wealth, $\Delta W_T = (1+r)s_{T-1}$, is just sufficient to set off the reduction of the young generation's contribution. In its own youth generation $T-1$ reduces its contribution by more than generation T because it has to finance the increase of its savings in addition to the tax payment. This is, again, anticipated by the preceding generation $T-2$, who has to increase savings by more than generation $T-1$ in order to set off this higher contribution reduction. Thinking on in this way we understand why the changes of contributions and savings are the higher, the closer we come to the transfer period t. In the first post-transfer period, $t+1$, the contribution reduction of the young generation reaches its maximum value, $(1+r)D$. Anticipating this, the generation who receives the transfer D will not increase consumption but save the entire transfer in order to set off the diminished contribution of its successors in the next period. In this way consumption levels and total public good supply remain unaffected by the policy change. The same holds for the utility levels of all generations, including the generation given the transfer.

From these considerations we can draw our central conclusion, which is summarized in the following proposition:

Proposition 2 *Assume that in the time interval* $[1, T]$ *both generations make permanently strictly positive contributions to the public good. If in an arbitrary period* $t \in [1, T-1]$ *the young generation is granted a once-and-for-all debt financed transfer, while the succeeding young generations in* $t+1, \ldots, T$ *are charged with a lump-sum tax in order to pay for the debt services, this will entail an intertemporal redistribution of incomes which is, however, neutral with respect to private and public consumption as well as to generations' welfare positions.*

5 Discussion and Conclusions

Which conclusions can be drawn from our neutrality result? First of all, it is important to notice that it has been derived under the assumption of a special utility function. Thus, we have to be cautious not to overemphasize its range of validity. However, we are quite confident that the result will be preserved for general utility functions. To show this will be a task of future research.

The question of generality left aside, Proposition 2 implies that the neutrality theorem derived by Warr (1983) in the static standard model of voluntary public good provision can be extended to the dynamic context of our model. To the same degree as the allocation of resources is unaffected by an *intra*temporal redistribution of incomes between contributors in the static framework, it is invariant with respect to an *inter*temporal redistribution of incomes between generations in our dynamic framework. A closer look at the causes for neutrality shows that this conclusion is less surprising than it might seem a priori.

In the standard model of public good provision the ineffectiveness of income redistributions can be seen quite easily. If we formulate, according to Cornes and Sandler (1984), the decision problems of the contributors in an appropriate way it is directly revealed that the agents calculate their optimal decisions on the basis of the *economy's total resource constraint*. As for our model, we have seen that the cause for intertemporal neutrality is very similar. When choosing their present private and public consumption levels, generations must take into account the strategic effects of savings on future private and public consumption levels. For that purpose they have to forecast the choices of all future generations, which requires the method of backward induction. After all, this procedure amounts to generations maximizing their utilities subject to an *intertemporal budget constraint which contains not only their own periodical incomes, but also the incomes accruing to all future generations*.[11] If capital markets are perfect, this budget constraint can never be altered by an (*intertemporal or intratemporal*) income redistribution. Therefore, there cannot be any real effects as long as generations are not forced into corner solutions, which would make it impossible for them to neutralize the income redistribution by adjustments of savings and contributions to the public good. Given the ineffectiveness of an intertemporal income redistribution we can conclude immediately that in our model the Ricardian Equivalence Theorem is valid. In fact, Ricardian Equivalence and the intertemporal version of Warr's neutrality theorem as derived in this paper coincide. This becomes obvious if we compare the way in which the public debt policy is neutralized in our model with the logic underlying the argumentation in Barro (1974). In short, this argumentation can be summarized as follows:

If individuals care about the utilities of their direct descendants, a debt financed transfer has no real effects, even if the time interval in which individuals have to pay taxes to finance the debt services exceeds their remaining lifetime. Although the present value of their disposable incomes is increased in this case, they will not increase their consumption levels. Instead, being altruistic, they will leave the entire transfer as an additional bequest to their children. This change of the bequest is just high enough to finance the future tax burden of the children and the additional bequests left by the children to the grandchildren. Since all individuals act in this way, their wealth is neither augmented nor reduced. Therefore, they will choose the same consumption pattern as in the absence of the debt policy. Thus, the policy does neither affect the allocation of resources, nor the individuals' welfare positions. As described above, exactly the same behavior of generations is observable in our public good model. Here, however, the neutralization of public debt is brought about by changes of the contributions to the public good rather than by altruistically motivated bequests. In this respect the cause for neutrality differs quite a lot from that in the Barro model.

There has been much discussion about the assumptions which cause Ricardian Equivalence. One of these assumptions is that the individuals' planning horizon

[11] Cf. Eqs. (6) and (19) or in more detail (Cornes and Hartley 2007, p. 212).

must contain the entire policy interval, including the transfer period and all periods in which taxes to finance the debt services have to be collected, cf. Brennan and Buchanan (1980, p. 5). If the duration of the debt policy exceeds individual lifetimes, a planning horizon long enough to bring about Ricardian Equivalence requires some kind of linkage between successive generations.

Barro (1974) showed that such an intergenerational linkage can be established by a bequest motive arising from utility related altruism. The central point in Barro's modeling of intergenerational linkages is that it connects the single generations to one dynastic family. The behavior of this dynasty can be equivalently described by a single long-lived individual who decides about the uses of the economy's total intertemporal resources. It is clear that, under these conditions, public debt must be neutral since it only leads to a different timing of the income stream without altering its present value. Thus, the dynasty's real opportunities are preserved.

Since Barro (1974) the question whether this kind of modeling intergenerational linkages is necessary and/or sufficient to bring about Ricardian Equivalence has played a central role in the neutrality debate. The results emanating from this debate are quite disturbing. In an important contribution Bernheim and Bagwell (1988) showed that altruistically motivated transfers lead to debt neutrality under far less restrictive assumptions on intergenerational linkages than those made by Barro. In particular, they proved that the assumption of the 'dynastic family' is not necessary for Ricardian Equivalence. Instead, Bernheim and Bagwell (1988, p. 313) stressed that it is sufficient that "…linkages are operative (in the sense that transfers are positive and equilibria are robust with respect to perturbations of corner constraints) …". In other words, debt neutrality hinges not upon the specific nature of the intergenerational linkage, but rather on the untouchedness of agents' preferred alternatives.[12] As a consequence, Bernheim and Bagwell (1988) identified a profusion of possible family linkages which lead to Ricardian Equivalence. The results of our paper suggest, that the intertemporal provision of a public good can apparently provide such an 'operative link' in the sense of Bernheim and Bagwell (1988). Again, the decisive point here is that generations calculate with the economy's total wealth as their disposable income, which is not affected by public debt. Note that this is true despite the fact that generations play non-cooperatively and can therefore certainly not be viewed as one dynastic family.

In the light of this discussion the question which remains to be answered is: Which kinds of intergenerational linkages are operative? Bernheim and Bagwell (1988, p. 334) concede that their comprehensive neutrality result depends critically on the assumption of altruism which is 'pure' in the sense that individuals derive no direct utility from giving. The case of 'impure' altruism was investigated by Andreoni (1989). Assuming that individuals care about the consumption of their heir and that, in addition, they get a 'warm glow' from the mere act of transferring income, he found that a redistributive policy is not neutral. Similar results are derived by Abel and Bernheim (1992) and Kotlikoff et al. (1990),

[12] A similar observation was made by Carmichael (1982).

who consider other types of models in which altruism is not 'frictionless'. In our model, generations are linked by the consumption of a public good rather than by intergenerational transfers. Moreover, the behavioral assumptions are far from being altruistic. Nevertheless, the Ricardian Equivalence Theorem is valid because the intergenerational linkage is operative. Thus, the most important conclusion which can be drawn from the results of this paper is, that debt neutrality is possible even in the absence of altruism. Thus, while the exact nature of intergenerational linkages required to generate Ricardian Equivalence remains obscure, it is clear that altruism is not a necessary condition.

Finally, as Andreoni (1989) suggested, the linkage between two generations requires some kind of a public good. In most models dealing with Ricardian Equivalence, this is done by parents looking for their descendants' consumption through strategically adjusting bequests. In a model with two generations caring for a public good, the intertemporal linkage works through strategically adjusting savings while being young.

References

Abel, A. B., & Bernheim, B. D. (1992). Fiscal policy with impure intergenerational altruism. *Econometrica, 59*, 1687–1711.

Andreoni, J. (1988). Privately provided public goods in a large economy: The limits of altruism. *Journal of Public Economics, 35*, 57–73.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian Equivalence. *Journal of Political Economy, 97*, 1447–1458.

Andreoni, J., & Bergstrom, T. (1996). Do government subsidies increase the supply of public goods? *Public Choice, 88*, 295–308.

Barro, R. J. (1974). Are government bonds net wealth? *Journal of Political Economy, 82*, 1095–1117.

Becker, G. S. (1974). A theory of social interactions. *Journal of Political Economy, 82*, 1063–1093.

Bergstrom, T., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics, 29*, 25–49.

Bernheim, B. D. (1986). On the voluntary and involuntary provision of public goods. *American Economic Review, 76*, 789–793.

Bernheim, B. D., & Bagwell, K. (1988). Is everything neutral? *Journal of Political Economy, 96*, 308–338.

Boadway, R., Pestieau, P., & Wildasin, D. (1989). Tax-transfer policies and the voluntary provision of public goods. *Journal of Public Economics, 39*, 157–176.

Brennan, G., & Buchanan, J. M. (1980). The logic of the Ricardian Equivalence theorem. *Finanzarchiv, 38*, 4–16.

Buchholz, W., Cornes, R. C., & Peters, W. (2006). On the frequency of interior Cournot Nash equilibria in a public good economy. *Journal of Public Economic Theory, 8*, 401–408.

Carmichael, J. (1982). On Barro's theorem of debt neutrality: The irrelevance of net wealth. *American Economic Review, 72*, 202–213.

Cornes, R. C., & Hartley, R. (2007). Aggregative public good games. *Journal of Public Economic Theory, 9*, 201–219.

Cornes, R. C., & Sandler, T. (1984). Easy riders, joint production, and public goods. *Economic Journal, 94*, 580–598.

Kotlikoff , L., Persson, T., & Svensson, L. (1988). Social contracts as assets: A possible solution to the time-consistency problem. *American Economic Review, 78*, 662–667.

Kotlikoff , L., Razin, A., & Rosenthal, R. W. (1990). A strategic altruism model in which Ricardian Equivalence does not hold. *Economic Journal, 100*, 1261–1268.

Nett, L., & Peters, W. (1993). Financing public goods: Voluntary contributions and income taxation. *Finanzarchiv, 50*, 515–529.

Rebelein, R. P. (2002). Ricardian equivalence survives strategic behavior. *Public Finance/Finances Publiques, 53*, 148–228.

Rebelein, R. P. (2006). Strategic behavior, debt neutrality and crowding out. *Public Finance Review, 34*, 148–172.

Warr, P. (1983). The private provision of public goods is independent of the distribution of incomes. *Economics Letters, 13*, 207–211.

Counterterrorism: A Public Goods Approach

Todd Sandler

1 Introduction

Unfortunately, terrorism occupies an increasing presence in the world with many ghastly events in recent years. Noteworthy attacks include al-Shabaab's attack on Westgate shopping mall in Nairobi, Kenya, on 21 September 2013; Islamic State in Iraq and Syria's (ISIS's) beheadings of hostages beginning in 2014; Boko Haram's kidnapping of 276 schoolgirls from Chibok, Nigeria, on 14-15 April 2014; ISIS's downing of Metrojet flight 9268 en route from Sharam el-Sheikh to St. Petersburg on 31 October 2015; ISIS's armed attacks in Paris at multiple venues on 13 November 2015; and ISIS's suicide bombings in Brussels at the airport and metro station on 22 March 2016. Other noteworthy past terrorism incidents include the suicide truck bombing of the US Marine barracks in Beirut on 23 October 1983; the downing of Air-India Boeing 747 en route from Montreal to London on 23 June 1985; the downing of Pan Am flight 103 en route from London to New York on 21 December 1988; the barricade hostage seizure of Moscow Theater by Chechen rebels on 23 October 2002; and the bombing of commuter trains and station in Madrid on 11 March 2004.

Academic interest in the study of counterterrorism had its roots in the late 1960s at the beginning of the era of transnational terrorism when terrorist attacks had implications for two or more countries (Enders and Sandler 2012; Hoffman 2006). Terrorists resorted to transnational terrorist attacks in order to turn the world's attention to their cause. At first, transnational terrorist groups, such as the Popular Front for the Liberation of Palestine (PFLP), skyjacked commercial flights en route to foreign designations because satellite transmission of the event made everyone

T. Sandler (✉)
Department of Economics, School of Economic, Political and Policy Sciences, University of Texas at Dallas, 800 W. Campbell Rd., Richardson, TX, 75080 USA
e-mail: tsandler@utdallas.edu

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_11

acutely aware of them and their cause. In addition, terrorist groups sent squads to foreign capital cities, where their attack captured the world's attention—e.g., the abduction of Israeli's athletes at the 1972 Munich Olympics. Recent empirical work on homegrown and home-directed domestic terrorism showed that domestic terrorist groups graduated to transnational terrorist incidents when home campaigns generated little notice (Enders et al. 2011).

Academic analysis of terrorism grew greatly after al-Qaida's four hijackings in the United States on 11 September 2001 (henceforth, 9/11) that resulted in the collapse of the World Trade Center towers, partial destruction of the Pentagon, and a plane crash in Shanksville, Pennsylvania. In total, almost 3000 people perished in these hijackings and many were injured. Subsequent studies applied sophisticated theoretical and empirical methods to the study of terrorism and counterterrorism (Sandler 2014). The former explored the root causes of terrorism and the role played by income (Bandyopadhyay and Younas 2011; Enders et al. 2016; Gassebner and Luechinger 2011; Krueger and Maleckova 2003), globalization (Dreher et al. 2008; Li and Schaub 2004), regime type (Eubank and Weinberg 1994; Eyerman 1998; Gaibulloev et al. 2017; Piazza 2007, 2008; Sandler 1995), and other grievance-causing factors (e.g., economic discrimination). The theoretical study of counterterrorism involves the application of game theory to investigate the interaction among targeted governments or the interface between a terrorist group and one or more targeted governments (Bandyopadhyay and Sandler 2011; Cárceles-Poveda and Tauman 2011; Sandler and Lapan 1988; Sandler and Siqueira 2006; Schneider et al. 2015; Siqueira 2005; Siqueira and Sandler 2006). Noncooperative game theory is an ideal tool to analyze counterterrorism because adversaries or allies are taking independent actions to further their self-interest subject to their constraints and the anticipated response of their counterparts. A government's countermeasures affect a terrorist group's constraint, while terrorist attacks influence a government's objective or constraints. Thus, targeted countries fortify their borders in the hopes of deflecting terrorist attacks to alternative less fortified countries (Gardeazabal and Sandler 2015). Countries targeted by the same terrorist group may do little in the hopes that other attacked countries will take offensive measures to weaken the common terrorist threat. In short, game theory casts the analysis of counterterrorism into one involving strategic rational choice on the part of the agents.

In its game-theoretic formulation, the study of counterterrorism concerns myriad concepts of public goods and externalities. The purpose of this chapter is to underscore the broad-ranging contributions of Richard Cornes and his co-authors by demonstrating how their methods provide a theoretical foundation for better understanding the practice of counterterrorism. To do so, I apply aspects of the private provision of public good model (Bergstrom et al. 1986; Cornes and Sandler 1985, 1986, 1996; Cornes et al. 1999). Throughout the ensuing chapter, I employ the Cornes and Sandler (1984, 1985) graphical device to elucidate numerous insights about counterterrorism in various scenarios. In the case of intelligence gathering, the joint product model plays a role where a single activity yields multiple outputs that vary in their degree of publicness (Cornes and Sandler 1984, 1994).

Public goods and externalities are tied to the two primary types of counterterrorism: proactive and defensive measures. The former are offensive actions intended to reduce the assets or capabilities of the terrorist group. Such actions may involve assassinating terrorist leaders, capturing terrorist operatives, reducing terrorist finances, infiltrating terrorist groups, or gathering intelligence. Effective proactive responses by any targeted country curb the threat for all at-risk countries, thereby providing a public good. In contrast, defensive countermeasures make it more difficult for terrorists to attack successfully. In the event of a terrorist attack, defensive actions limit the resulting damage or loss of lives. As shown later, defensive counterterrorism generates a complex mix of externalities. Even the interaction between terrorists and governments are better understood with methods developed by Cornes and co-authors.

The remainder of the chapter has six sections. Preliminaries concerning the methods applied and the notion of terrorism are presented in Sect. 2. Proactive measures are analyzed in Sect. 3, followed by an investigation of defensive responses in Sect. 4. Section 5 examines the publicness of intelligence gathering. The interplay between a government and a terrorist group is addressed in Sect. 6. Finally, Sect. 7 contains concluding remarks.

2 Preliminaries

Much of the chapter focuses on two-player games where player i chooses to minimize cost, $C_i(q_i, q_j)$, or maximize utility, $U_i(q_i, q_j)$, subject to constraints that include a fixed parameter, q_j, representing player j's choice variable. An analogous problem applies to player j. The respective agents' choice variables are continuously differentiable and often denote counterterrorism actions in the ensuing study. For illustration, we express a few essential definitions in terms of a cost-minimization problem. Player i's best response, BR_i, to agent j's choice, q_j, is

$$q_i = BR_i(q_j) = \arg\min_{q_i} C_i\left(q_i, q_j\right), \tag{1}$$

while player j's best response, BR_j, to agent i's choice q_i, is

$$q_j = BR_j(q_i) = \arg\min_{q_j} C_j\left(q_i, q_j\right). \tag{2}$$

The best-response function for agent i is found by solving $\partial C_i\left(q_i, q_j\right)/\partial q_i = 0$ implicitly for q_i in terms of q_j. Analogously, the best-response function for agent

j is found by solving $\partial C_j\left(q_i, q_j\right)/\partial q_j = 0$ implicitly for q_j in terms of q_i. The simultaneous solution to (1)–(2) gives:

Definition 1 Strategy profile (q_i^N, q_j^N) is a Nash equilibrium if and only if $q_i^N \in \arg\min_{q_i} C_i\left(q_i, q_j^N\right)$ and $q_j^N \in \arg\min_{q_j} C_j\left(q_i^N, q_j\right)$.

At the Nash equilibrium, each agent's choice is a best response to that of the other agent, leaving neither agent to want to change unilaterally its decision variable if offered the opportunity to do so. The problem may include n agents by replacing q_j by $q_{-i} = (q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_n)$ and requiring that Definition 1 holds for q_i^N for $i = 1, \ldots, n$. The analysis herein will generally stay with the two-agent case. In places, an additive technology—i.e., $q_i + q_j$—will be applied, indicative of pure public goods and less indicative of general externalities (Cornes and Sandler 1996).

Two other crucial definitions for the two-agent case are:

Definition 2 Strategies q_i and q_j are strategic substitutes if the slopes of the best-response curves are negative—i.e., $\partial BR_i/\partial q_j < 0$ and $\partial BR_j/\partial q_i < 0$; and

Definition 3 Strategies q_i and q_j are strategic complements if the slopes of the best-response curves are positive—i.e., $\partial BR_i/\partial q_j > 0$ and $\partial BR_j/\partial q_i > 0$.

These definitions are due to Bulow et al. (1985). Strategic substitutes indicate that one agent's action replaces the need for the other agent's action, whereas strategic complements mean that one agent's action encourages this action on the part of the other agent. Generally speaking, contributions to a pure public good represent strategic substitutes, while exploitation efforts in an open-access commons represent strategic complements. In the latter case, harvesting efforts result in a race to exploit the openly available common property resource (Cornes and Sandler 1983). Arms race constitutes another instance of strategic complement, where one adversary's buildup of forces induces further buildup by the rival country.

Another important notion is that of plain substitutes and plain complements, as defined by Eaton (2004) and Eaton and Eswaran (2002). These concepts correspond to the cross-partials of the objective function, unlike strategic substitutes and complements which correspond to the cross-partials of the marginal function, as previously expressed in Definitions 2 and 3. For plain complements and cost minimization, $\partial C_i/\partial q_j < 0$, so that increased effort by one's counterpart reduces one own's cost, which is a good thing. In contrast, plain substitutes involve $\partial C_i/\partial q_j > 0$, or greater costs resulting from the actions of one's counterpart. For maximizing utility or profit, $\partial U_i/\partial q_j > 0$ denotes a plain complement, while $\partial U_i/\partial q_j < 0$ indicates a plain substitute.

Next, I turn to background on terrorism. Terrorism is the premeditated use or the threat to use violence by individuals or subnational groups to obtain a political objective through the intimidation of a large audience beyond that of the immediate victim (Enders and Sandler 2012). An essential ingredient of terrorism is the political objective, without which a kidnapping is an act of extortion and a bombing is a criminal act. The wider audience, which is typically a political constituency, is needed to pressure a government to concede to a terrorist group's political demands. Terrorism can be further subdivided into domestic and transnational terrorism (Enders et al. 2011). Domestic terrorism is homegrown where the perpetrator and victims are citizens from the venue country of the attack. Generally, a central government can internalize the externalities that terrorism in different provinces or locations in the same country implies. Transnational terrorism involves two or more countries owing to the nationalities of the perpetrators or victims in regards to the venue country. A skyjacking that originates in one country and concludes in another country is an instance of transnational terrorism. The beheadings of Western hostages by ISIS terrorists are acts of transnational terrorism, as are armed attacks by terrorists in a foreign capital. This chapter focuses on transnational terrorism, which involves the presence of transnational externalities from the practice of counterterrorism. Unless countries cooperate with one another, these externalities will not be internalized. Proactive countermeasures often result in underprovision compared to a Pareto-efficient ideal, while defensive action may imply overprovision or underprovision depending on the mix of externalities.

3 Proactive Measures

Consider a scenario where countries i and j are targeted by the same terrorist network that can strike at each country's interests at home and abroad. The common terrorist threat confronting the two countries can be reduced through proactive measures (e.g., drone attacks against the terrorist network's assets). Such measures have strong elements of publicness—i.e., nonexcludability and nonrivalry of benefits. Without loss of generality, I examine country i's viewpoint since the equations are symmetric for country j. Proactive measure, q_i, results in three cost and benefit components: the cost of the action, potential losses to i from a terrorist attack at home, and potential losses from a terrorist attack on i's interests in country j. Proactive cost is denoted by $G(q_i)$ with $G'(q_i) > 0$ and $G''(q_i) > 0$, so that this cost increases at an increasing rate. The expected loss from a home attack on i is $\pi_i l(q_i)$ where $l'(q_i) < 0$, so that i's proactive measures reduces its potential losses by weakening the terrorists' ability to inflict harm in country i on i's home interests.[1] The likelihood of an attack in country i is $\pi_i(q_i, q_j)$ with $\partial\pi_i/\partial q_i < 0$, $\partial\pi_i/\partial q_j < 0$, and $\partial^2\pi_i/\partial q_i \partial q_j > 0$. Proactive measures by either country reduce the likelihood of an attack on i as the terrorists' capabilities are weakened. The cross-partial of π_i is positive, indicative of substitutes and diminishing returns to effort. In addition, i's proactive response limits its expected losses in j, which is denoted by $\pi_j v(q_j)$,

[1] In a more general model, we could write i's loss as $l(q_i, q_j)$, so that j's proactive efforts also limit i's losses at home. This would provide more publicness and externalities. Because the likelihood of attack depends on both countries' proactive measures, the model has plenty of externalities without this further complication.

where $\pi_j(q_i, q_j)$ has negative first-order partials and positive cross-partial analogous to π_i. Country i's losses in country j are denoted by $v(q_j)$ with $v'(q_j) < 0$, so that these losses are limited by j's proactive response. Offensive measures by i protects its interests abroad by reducing the likelihood of an attack abroad.

Country i chooses its proactive level to minimize its cost, C_i, as follows:

$$\min_{q_i} C_i = \left[G(q_i) + \pi_i\left(q_i, q_j\right) l(q_i) + \pi_j\left(q_i, q_j\right) v(q_j)\right], \tag{3}$$

where the second term is the expected cost of an attack in country i on its home interests and the third term is the expected cost on i's interests of an attack in country j. The associated first-order condition (FOC) is:

$$\frac{\partial C_i}{\partial q_i} = G'(q_i) + l(q_i)\frac{\partial \pi_i}{\partial q_i} + \pi_i l'(q_i) + v(q_j)\frac{\partial \pi_j}{\partial q_i} = 0, \tag{4}$$

with second-order condition (SOC), $\partial^2 C_i/\partial q_i^2 > 0$ or strict convexity of cost function.[2] In (4), there are three marginal benefits and one marginal cost. By increasing its offensive measures, country i reduces not only the likelihood of a home attack, but also the damage of a home attack. This corresponds to the second and third right-hand terms of (4), which are negative marginal losses or marginal benefits. This proactive measure also decreases the likelihood of an attack in country j, thereby guarding i's foreign assets. Thus, the fourth right-hand term in (4) is also a marginal benefit. The three marginal benefits are traded off against the increased cost of taking such action, which is the first right-hand term in (4). An analogous objective and FOC hold for country j and are not displayed. A mere switch of the i and j subscripts is required. Equation (4) implicitly indicates i's best response, $BR_i = q_i$, in terms of q_j. To establish that this is a case of strategic substitutes, I apply the implicit function rule to (4) to get:

$$\frac{\partial BR_i}{\partial q_j} = \frac{-l'(q_i)\frac{\partial \pi_i}{\partial q_j} - v'(q_j)\frac{\partial \pi_j}{\partial q_i} - l(q_i)\frac{\partial^2 \pi_i}{\partial q_i \partial q_j} - v(q_j)\frac{\partial^2 \pi_j}{\partial q_i \partial q_j}}{\partial^2 C_i/\partial q_i^2} < 0. \tag{5}$$

The negative sign, indicative of strategic substitutes, follows because all four terms in the numerator are negative and the denominator is positive by the SOC. Similarly, we can show that $\partial BR_j/\partial q_i < 0$ for j's best-response function.

The Cornes-Sandler diagram can now be instructive for this case of strategic substitutes. In Fig. 1, I revert to $i, j = 1, 2$ with q_1 on the horizontal axis and q_2 on the vertical axis. IC_1 represents country 1's isocost curve for alternative proactive responses for the two countries, which corresponds to a constant level of C_1 in the bracketed expression in (3). IC_1's U-shaped contour follows because small q_1

[2] Throughout the analysis, I ignore corner solutions where all attacks are avoided, $\pi_i + \pi_j = 0$, or where only one country takes offensive measures.

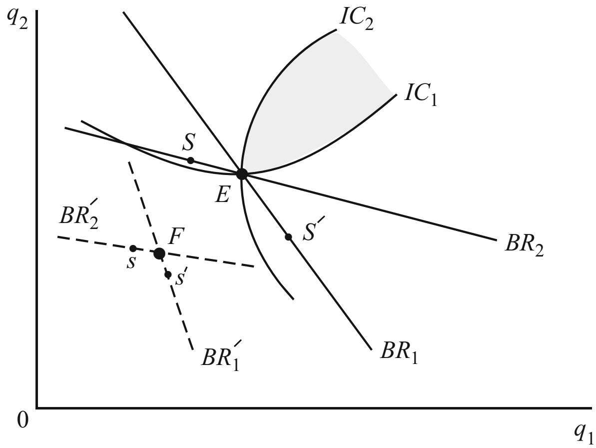


Fig. 1 Proactive counterterrorism measures

is associated with marginal proactive benefits overwhelming marginal proactive cost, while large q_1 is associated with marginal proactive cost exceeding marginal proactive benefits.[3] The minima of IC_1 satisfies (4) and is on the Nash reaction path or best-response curve, BR_1, of country 1. Partial differentiation of C_1 with respect to q_2 gives

$$\frac{\partial C_1}{\partial q_2} = l(q_1)\frac{\partial \pi_1}{\partial q_2} + v(q_2)\frac{\partial \pi_2}{\partial q_2} + \pi_2 v'(q_2) < 0. \tag{6}$$

This expression tells us that the area above country 1's isocost curve denotes reduced cost for country 1 for greater q_2. This, in turn, implies plain complements, where 2's proactive effort increases 1's well-being. In Fig. 1, IC_2 represents one of country 2's C-shaped isocost curves, for which the area to the east of a given curve corresponds to reduced cost and, thus, greater well-being. Equation (5), tailored to country 2, indicates that 2's best-response path, BR_2, is also negatively sloped, indicative of strategic substitutes. BR_2 connects infinite-sloped points on the IC_2 contours where 2's FOC is satisfied. For simplicity, the two best-response paths (BR_1 and BR_2) are drawn in a linear fashion; however, these paths may be curvilinear. To satisfy stability and uniqueness of equilibrium, the slope of BR_1 must be less than –1 and greater than $-\infty$ and the slope of BR_2 must be greater than –1 and less than zero (Cornes et al. 1999; Cornes and Sandler 1996).

In Fig. 1, the Nash equilibrium, E, occurs at the intersection of the two best-response paths where the slope of IC_1 is zero and that of IC_2 is infinite. At this intersection, both countries satisfy their FOC and, hence, Definition 1 holds. The

[3]The slope of IC_1 is found by the implicit function rule applied to C_1 to give an expression for $\partial q_2/\partial q_1$ for a constant C_1. The numerator of this partial derivative is the FOC in (4). The second-order partial is positive.

shaded lens-shaped region denotes Pareto superior points to E, where both countries experience smaller cost. The total provision of proactive efforts at E can be found by drawing a line with slope –1 from E to the q_1 axis (not shown). Drawing a similar line from any point in the shaded area results in greater provision of proactive measures. Thus, the Nash equilibrium implies underprovision.

This underprovision can be shown rigorously by first finding the minimization of total cost, C^T, for the two targeted countries:

$$\min_{q_i, q_j} C^T = G(q_i) + G(q_j) + \pi_i(\mathbf{q})\left[l(q_i) + v(q_i)\right] + \pi_j(\mathbf{q})\left[l(q_j) + v(q_j)\right], \quad (7)$$

where $\mathbf{q} = (q_i, q_j)$. The FOC for country i is:

$$\frac{\partial C^T}{\partial q_i} = G'(q_i) + \pi_i\left[l'(q_i) + v'(q_i)\right] + \left[l(q_i) + v(q_i)\right]\frac{\partial \pi_i}{\partial q_i} + \left[l(q_j) + v(q_j)\right]\frac{\partial \pi_j}{\partial q_i} = 0. \quad (8)$$

A similar expression holds for $\partial C^T/\partial q_j$. Evaluation of the FOC in (8) at the Nash equilibrium, $\mathbf{q^N} = \left(q_i^N, q_j^N\right)$, which satisfies (4), yields:

$$\pi_i v'\left(q_i^N\right) + v\left(q_i^N\right)\frac{\partial \pi_i}{\partial q_i} + l\left(q_j^N\right)\frac{\partial \pi_j}{\partial q_i} < 0. \quad (9)$$

Equation (9) follows because four expressions on the right-hand side of (8) must sum to zero when evaluated at the Nash equilibrium, thereby leaving just three terms. Each corresponds to reduced marginal external cost, thus a marginal external benefit. The first term is the marginal external benefit conferred by i's proactive efforts on limiting j's losses in country i, while the second term is the marginal external benefit conferred by i's proaction on reducing the likelihood of home attacks that damage j's assets in country i. In (9), the third term is the marginal external benefit stemming from i's proactive measures reducing the likelihood of attacks on j's interests at home. The sign of the inequality indicates underprovision of proactive measures. A similar analysis and conclusion applies to country j's offensive efforts.

3.1 Leadership and Unilateral Action

Next, consider leadership scenarios where country 1 first chooses q_1. The follower acts like a Nash player and abides by its best-response path, BR_2. Country 1 then chooses its q_1 using BR_2 as a constraint so that it seeks the tangency of its highest isocost curve to BR_2 at point S in Fig. 1. To limit clutter, I do not display the tangent IC. At this leader-follower equilibrium, the leader shifts some of the proactive burden onto the follower as $q_2^S > q_2^N$ and $q_1^S < q_1^N$. By extending a 45° line with slope –1 from S to the horizontal axis, one sees that the leader-follower equilibrium implies less overall proactive provision than the Nash equilibrium.

Analogously, if country 2 is the leader, then S' is the leader-follower equilibrium where 2's isocost curve (not shown) is tangent to BR_1. Once again, one sees that this outcome results in a smaller provision level than the Nash equilibrium at point E. This leadership cannot address underprovision when both best-response paths are negatively sloped since the leader accounts for the follower's negative conjectural variation. If, however, the leader's proactive measures induce further action by the follower owing to a behavioral response of wanting to match the leader's proactive efforts, then leadership can reduce underprovision (Buchholz and Sandler 2017). This alternative scenario requires a positive conjecture or positively sloped best-response path for the follower.

In a seminal piece, Hoel (1991) showed that unilateral action in a pure public good situation (e.g., removal of pollution) does not benefit the agent taking this action. If, say, country 1 assumes an altruistic attitude and internalizes some of the external protection that it provides so as to shift BR_1 in the northeast direction, then the new equilibrium on BR_2 implies greater cost for country 1. Moreover, country 2's efforts are lowered, which works against country 1's intentions. The shortfall of proactive measures is difficult to address without a fundamental change in the underlying model.

3.2 Backlash

Backlash occurs when proactive measures result in further terrorist grievances, which induce more terrorist recruitment and attacks (Rosendorff and Sandler 2004). ISIS wants the United States and Europe to take aggressive actions against the group in Iraq and Syria, so that ISIS can solicit more recruits. Attacks against innocent Muslims in the United States and Europe after the Paris and San Bernardino attacks play into ISIS's plan to recruit more converts.

Let $B(q_i)$ denote backlash cost, which is added to the objective in (3). The new FOC will have an additional marginal cost to weigh against the three marginal benefits, thereby resulting in smaller q_i for each q_j. As a consequence, country 1's best-response path shifts down and becomes steeper—i.e., BR_1' in Fig. 1, while country 2's best-response path shifts down and becomes flatter—i.e., BR_2' in Fig. 1. The new Nash equilibrium is at F where the two dashed best-response paths intersect in Fig. 1 in the light of backlash. With backlash, there is less of a gain to leadership because of the enhanced cost of assuming an offensive stance by drawing an attack. The leader-follower equilibrium for backlash will be nearer to the Nash equilibrium (see points s and s' in Fig. 1). Backlash can also be captured through a *smaller ability* to reduce the likelihood of the attack through proactive effort. This latter consideration also makes BR_1 steeper and BR_2 flatter and shifts both paths down.

4 Defensive Measures

Next, I tailor the model to account for defensive measures, intended to reduce the likelihood of a terrorist attack and to limit the resulting damage of successful attacks. Defensive actions imply an interesting set of opposing externalities since a country must be cognizant of its interests at home and abroad. Greater defense at home may merely deflect an attack abroad, where its assets are targeted and the country must depend on the venue country to protect these assets. Although I limit these opposing externalities to a bare minimum, many additional externalities could be introduced.

The primary modeling difference concerns the likelihood of attack, $\pi_i(q_i, q_j)$, for which own defense reduces terrorist attacks at home, $\partial\pi_i/\partial q_i < 0$, but increases these attacks abroad, $\partial\pi_j/\partial q_i > 0$ for i and j and $i \neq j$. There is diminishing marginal returns to defensive efforts, $\partial^2\pi_i/\partial q_i^2 > 0$ and $\partial^2\pi_j/\partial q_i^2 < 0$. Moreover, country j's defense reduces (increases) the marginal impact of country i's action to limit terrorist attacks at home if j's defense is larger (smaller) than that of i, so that

$$\frac{\partial^2\pi_i}{\partial q_i \partial q_j} \gtreqless 0 \quad \text{and} \quad q_i \lesseqgtr q_j. \tag{10}$$

The likelihood of attack function is assumed symmetric between the two countries, such that $\pi_i\left(q_i, q_j\right) = \pi_j\left(q_j, q_i\right)$.

For defensive measures, the objective function is still (3) and the FOC is still (4); however, the fourth term in (4) is now a marginal cost as defense in country i deflects the attack to country j where i's foreign assets are jeopardized—i.e., $v(q_j)\left(\partial\pi_j/\partial q_i\right) > 0$. This change means that two marginal cost expressions are traded off against two marginal benefit terms. These benefits arise from greater safety stemming from home defense. The all-important slope of i's best-response curve has the same form as (5); however, its sign is now anticipated to be positive. At a symmetric solution where $q_i = q_j$, the cross partials, $\partial^2\pi_i/\partial q_i\partial q_j$ and $\partial^2\pi_j/\partial q_i\partial q_j$, in the numerator are zero.[4] The first two terms in the numerator are now positive because of attack transference (i.e., $\partial\pi_i/\partial q_j > 0$ and $\partial\pi_j/\partial q_i > 0$). Since the denominator is also positive, the best-response curves are now positively sloped in the neighborhood of the symmetric equilibrium, indicative of strategic complements. Even without the symmetry assumption, this is the likely outcome

[4]Without symmetry, these cross partials differ in sign so that the last two terms may offset to some degree. If q_i exceeds q_j, then the net difference is positive and reinforces the positivity of the numerator.

given that i's interests at home are generally greater than those abroad. Thus, defensive measures results in a fortification race, analogous to an arms race.

Evaluation of the FOC for the cooperative problem at the Nash equilibrium yields an inequality with the same terms as (9); however, this inequality cannot be signed owing to opposing externalities given each country's interests at home and abroad. Thus, more structure is required, which is provided by two special cases.

4.1 Host-Country Specific Assets

This case rules out each country having foreign assets, so that

$$v(q_i) = v'(q_i) = 0 \quad \text{for} \quad i = 1, 2. \tag{11}$$

The evaluation of the FOC that characterizes the cooperative solution at the Nash equilibrium implies,

$$\partial C^T/\partial q_i = l\left(q_j^N\right)\left(\partial \pi_j/\partial q_i\right) > 0, \quad i, j = 1, 2, \quad \text{and} \quad i \neq j, \tag{12}$$

so that independent behavior results in overprovision of defense as both countries ignores the transference externality that its fortification causes. Without foreign assets, there is no inhibiting factor that can attenuate the motive to deflect the attack abroad. If, say, country 1 has no foreign assets, while country 2 has assets in country 1, then country 1 will pursue transference to a greater extent than country 2.

In Fig. 2, the two solid best-response curves are displayed along with the Nash equilibrium at point E. Country 1's isocurve is now an inverted U-shaped contour

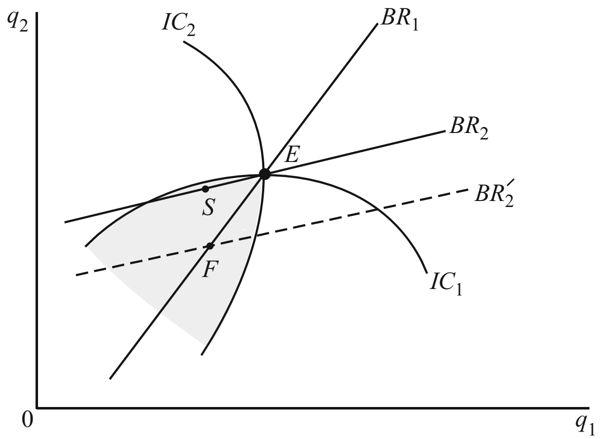


Fig. 2 Defensive race: no foreign assets

for which cost is smaller below the contour, given that

$$\frac{\partial C_i}{\partial q_j} = l(q_i)\frac{\partial \pi_i}{\partial q_j} > 0, \quad i,j = 1,2, \quad \text{and} \quad i \neq j. \tag{13}$$

Equation (13) indicates that i's cost increases with q_j as there is transference to country i. In Fig. 2, country 2's isocost contours are inverted C-shaped curves, where cost falls to the west of the contours. Equation (13) is consistent with plain substitutes. BR_1 joins the maxima of 1's isocost curves, while BR_2 joins the infinite-sloped points of 2's isocost curves. The intersection of these best-response paths results in the Nash equilibrium at E. The shaded lens-shaped region indicates Pareto improvement over the Nash equilibrium. Consistent with (12), these improvement points involve reduced defense on the part of the two countries.

This situation of strategic complements and plain substitutes implies a different outcome for leadership than was true for proactive measures. In Fig. 2, leadership by country 1 gives equilibrium S, at which 1's isocost curve is tangent to BR_2. S implies reduced levels of defensive measures by both countries. Because the reduction in q_1 is relatively greater than that of q_2, there is a second-mover advantage for the follower. Suppose that country 2 assumes unilateral action and account for some of the transference externality by reducing its defense for each level of q_1. This gives rise to the downward shift in BR_2 to the dashed $BR_2^{'}$ locus and the new equilibrium at F where both countries are better off. Ideally, unilateral behavior on the part of both countries could achieve a Pareto optimum provided that each internalizes the transference externality when deciding its defense. Obviously, some protection is required to limit consequences when attacks succeed, which may include first-responder resources among other things.

4.2 *Case 2: Globalized Threat*

This globalized threat scenario is where each country's losses from a terrorist attack are the same at home and abroad, so that

$$l(q_i) = v(q_j) \quad \text{and} \quad l(q_j) = v(q_i), \quad \text{if} \quad q_i = q_j. \tag{14}$$

As such, countries lose as much from a home attack as from an attack abroad. Globalization diffuses countries' assets worldwide and in the limit would achieve this scenario, which eliminates countries' incentive to divert attacks abroad. The countries' best-response paths are still positively sloped, indicating strategic complements. However, the isocost curves in Fig. 3 change their orientation as compared to the host-country-specific asset case. At a symmetric equilibrium where $q_i = q_j$ and $\partial\pi_i/\partial q_j = -\partial\pi_j/\partial q_j$, the following holds:

$$\frac{\partial C_i}{\partial q_j} = \pi_j v'(q_j) < 0 \quad \text{for} \quad i,j = 1,2, \quad \text{and} \quad i \neq j. \tag{15}$$

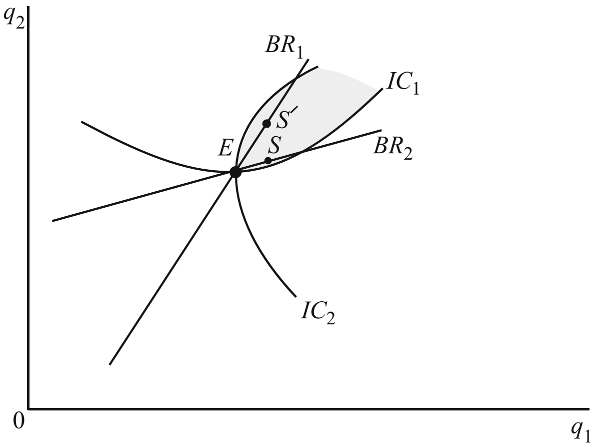


Fig. 3 Defensive countermeasures: generalized threat

Equation (15) indicates that the areas above IC_1 and to the east of IC_2 are lower (greater) levels of cost (well-being) for countries 1 and 2, respectively. Each country's defense protects both countries' interests on the defender's home soil. This justifies the U-shaped and C-shaped isocost contours for countries 1 and 2, respectively, in Fig. 3. If I were to distinguish between the loss functions within each country—i.e., $l_i(q_i) \neq l_j(q_j)$ and $v_i(q_i) \neq v_j(q_j)$, then a symmetric equilibrium would not be applicable. Nevertheless, the outcomes indicated next still hold.

The generalized or globalized threat case is one of strategic and plain complements. In Fig. 3, the Nash equilibrium is at E. In relation to E, Pareto improvement involves the shaded lens-shaped region, defined by IC_1 and IC_2 associated with the Nash equilibrium. An evaluation of C^T at the Nash equilibrium gives

$$\frac{\partial C^T}{\partial q_i} = \pi_i v'\left(q_i^N\right) < 0, \quad i = 1, 2, \tag{16}$$

which indicates underprovision of defense as external benefits that home protection affords foreign interests at home are not internalized. This supports the undersupply displayed in Fig. 3.

Unlike proactive measures, leadership can reduce this underprovision. In Fig. 3, leadership by country 1 results in outcome S, while leadership by country 2 results in outcome S'. Leader-follower behavior leads to larger aggregate provision of defense, which improves both countries' well-being by boosting defense. There appears to be a second-mover advantage as the follower increases its defense by a smaller amount than the leader, which is reflected by the relative shifts of the isocost curves. Unilateral effort in terms of a rightward shift of BR_1 or an upward shift of BR_2 improves general well-being *if not taken too far*.

This is not an aggregative game since there is no additive structure for q_i and q_j (Cornes and Hartley 2007). An aggregative game structure could be introduced for some proactive and defensive counterterrorism scenarios. Thus far, a two-country scenario is assumed. Increasing the number of countries results in a complex set of opposing and reinforcing externalities. Terrorist preferences and grievances result in some subset of countries facing no terrorist threat and, thus, no need for any counterterrorism measures. Corner solutions become relevant for these countries. In contrast, the United States confronts a huge terrorist threat [5] and is more motivated to take proactive measures (Bandyopadhyay and Sandler 2011). Without these measures to reduce terrorist capabilities, US defensive efforts must be large and never able to protect all potential targets.

5 Intelligence Gathering

The gathering of intelligence against a terrorist threat is a proactive measure that can be modeled in many ways, depending on how the two targeted countries' intelligence gathering adds to a country's knowledge of the terrorist threat.

5.1 Intelligence With Congestion

In this case, intelligence is analogous to a commons' crowding (and uncoordinated gathering) that jeopardizes each country's independent action in the field (Cornes and Sandler 1983). Consider a scenario where targeted countries are independently infiltrating a terrorist group. Given the reciprocal secrecy, each intelligence agent has no way of knowing that the other country's agent is not a terrorist in the infiltrated group. This failure of coordination may limit the true intelligence ascertained. In some instances, these independent actions may result in a tragic outcome where agents are accidently killed. Even without this extreme outcome, each country's agent jeopardizes the safety of the other country's agents in an infiltration situation.

Let x_i and x_j denote the level of intelligence gathering in the respective countries. Country i faces the following utility, U_i, optimizing problem:

$$\max_{x_i} U_i \left[w_i - px_i, I\left(x_i, x_j\right)\right], \tag{17}$$

where the budget constraint, $w_i = y_i + px_i$, is substituted for the private numéraire good, y_i, whose price is unity. U_i increases with greater levels of the private good and intelligence, I, but at a diminishing rate of increase.[6] The private good and i's intelligence are assumed to be Edgeworth complements, so that $\partial^2 U_i / \partial y_i \partial I > 0$.

[5] Almost 40 % of all transnational terrorist attacks were directed at US interests during 1968–2010 (Enders and Sandler 2012).

[6] To limit superscripts, I assign $\bar{I}$ to denote j's intelligence.

In (17), the country's unit cost of intelligence collecting is p. Given the crowding assumption, i's intelligence level rises with x_i and falls with x_j, so that $\partial I/\partial x_i = I_i > 0$ and $\partial I/\partial x_j = I_j < 0$. Moreover, crowding implies that the cross-partial I_{ij} is negative. In this first representation, intelligence is not a public good, given excludability due to secrecy and rivalry due to crowding.

The FOC for (17) is

$$-p\frac{\partial U_i}{\partial y_i} + I_i\frac{\partial U_i}{\partial I} = 0. \tag{18}$$

The corresponding slope of i's best-response path to changes in x_j is

$$\frac{\partial BR_i}{\partial x_j} = -\frac{\left[-p\frac{\partial^2 U_i}{\partial y_i \partial I}I_j + I_i I_j \frac{\partial^2 U_i}{\partial I^2} + I_{ij}\frac{\partial U_i}{\partial I}\right]}{\partial^2 U_i/\partial x_i^2} > 0, \tag{19}$$

where the SOC requires $\partial^2 U_i/\partial x_i^2 < 0$, which is a short-hand notation for the first derivative of the left-hand side of (18) which respect to x_i. Provided the first two terms in the numerator exceed the absolute value of the third term, Equation (19) indicates strategic complements, while

$$\frac{\partial U_i}{\partial x_j} = I_j\frac{\partial U_i}{\partial I} < 0 \tag{20}$$

is consistent with plain substitutes. The latter means that beneath an income-constrained isoutility contour, IU, i's utility increases as j's intelligence decreases. If IC_1 and IC_2 are relabeled as IU_1 and IU_2 and the axes are relabeled with x_2 and x_1, then Fig. 2 would serve to illustrate this intelligence congestion scenario.[7] As such, both leadership and unilateral action by either country improves both countries' well-being by recognizing the self-defeating intelligence race that ensues.

If, however, the countries were to share intelligence and jointly participate in such operations, then the form of the intelligence function is $I\left(x_i + x_j\right)$. As a consequence, the model corresponds to the private provision of a pure public good. This sharing ends targeted countries working at cross-purposes, but it does not ensure optimality because of free-riding incentives. Countries sometimes share intelligence on terrorist groups, but are loath to conduct joint intelligence operations; hence, the crowding scenario is generally appropriate.

[7] Implicit differentiation of the constrained utility function in (17) gives the slope of these isoutility functions, denoted by $\partial x_2/\partial x_1$. The partial derivative of this slope with respect to x_1 establishes the inverted U-shape to the contours. Similarly, 2's isoutility contours are inverted C-shaped curves.

5.2 Joint Products and Intelligence

The crowding representation of intelligence can be broadened to allow country-specific benefits and purely public crowding externalities. The latter is denoted by $\widehat{C}\left(x_i + x_j\right)$, where $X = x_i + x_j$, $d\widehat{C}/dX > 0$, and $d^2\widehat{C}/dX^2 > 0$. The income-constrained utility objective is

$$\max_{x_i} U_i \left[w_i - px_i, x_i, \widehat{C}(X) \right], \tag{21}$$

where $\partial U_i/\partial y_i > 0$, $\partial U_i/\partial x_i > 0$, and $\partial U_i/\partial \widehat{C} < 0$. Marginal utility for y_i and x_i diminishes; marginal disutility for crowding also diminishes $\left(\partial^2 U_i/\partial \widehat{C}^2 < 0\right)$. Finally, I assume $\partial^2 U_i/\partial y_i \partial \widehat{C} < 0$ and $\partial^2 U_i/\partial x_i \partial \widehat{C} < 0$.

The FOC is

$$-p\frac{\partial U_i}{\partial y_i} + \frac{\partial U_i}{\partial x_i} + \frac{\partial U_i}{\partial \widehat{C}}\frac{d\widehat{C}}{dX} = 0, \tag{22}$$

for which $\partial U_i/\partial x_i$ must equal the sum of the other two negative terms (Cornes and Sandler 1984, 1996).[8]

Implicit differentiation of (22) gives

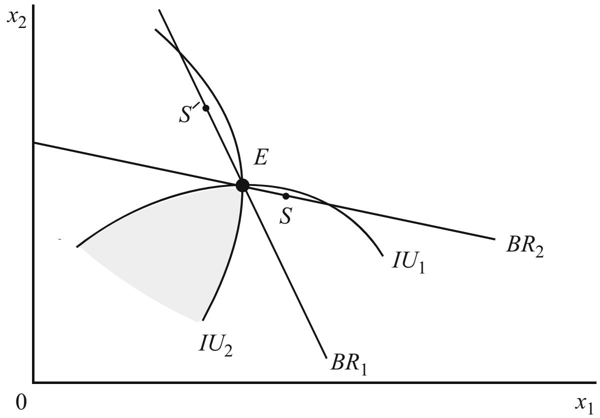
$$\frac{\partial BR_i}{\partial x_j} = -\frac{\left(-p\frac{\partial^2 U_i}{\partial y_i \partial \widehat{C}}\frac{d\widehat{C}}{dX} + \frac{\partial^2 U_i}{\partial x_i \partial \widehat{C}}\frac{d\widehat{C}}{dX} + \frac{\partial^2 U_i}{\partial \widehat{C}^2}\left(\frac{d\widehat{C}}{dX}\right)^2 + \frac{\partial U_i}{\partial \widehat{C}}\frac{d^2\widehat{C}}{dX^2}\right)}{\partial^2 U_i/\partial x_i^2} < 0, \tag{23}$$

where $\partial^2 U_i/\partial x_i^2 < 0$ is again a short-hand expression for the SOC, which is negative. Every term in the numerator is negative, consistent with strategic substitutes and a negatively sloped best-response path. The analogous expression, $\partial BR_j/\partial x_i$, is also negative, so that this is a situation of strategic substitutes. This is also an instance of plain substitutes, since $\partial U_i/\partial x_j = \left(\partial U_i/\partial \widehat{C}\right)\left(d\widehat{C}/dX\right) < 0$ for i and $j = 1, 2$ and $i \neq j$; hence i's (j's) isoutility curves are inverted U-shaped (C-shaped) contours.

This intelligence scenario is displayed in Fig. 4, along with the isoutility contours (IU_1 and IU_2) associated with the Nash equilibrium, E. The shaded lens-shaped area denotes Pareto superior points relative to E. Leadership by either country increases its intelligence operation relative to that of its counterpart (see S and S'), thereby improving its well-being at the expense of the other country. The same is true of unilateral action (not shown) that increases BR_1 and BR_2 provided the shifts are not too far—e.g., a shift in BR_1 to an intersection beyond where IU_1 crosses BR_2.

[8] A characteristic approach can be used to investigate this case—see Cornes and Sandler (1994).

Fig. 4 Intelligence provision



The two cases of intelligence congestion are interesting because the first implies strategic complements and the second implies strategic substitutes. As a consequence, one instance gains from leadership and unilateral action, while the other selectively gains from leadership and unilateral action. Thus, even slight modeling changes can drastically influence strategic and policy aspects for congestion-based models. This lesson can be applied to a host of nonterrorism scenarios, such as congestion-based tolls for highways where the form of the congestion function assumes an essential role (Cornes and Sandler 1996).

Thus far, asymmetric information has not been built into the above representations. This can be done and the terrorist group can be brought in as active informed agent (Arce and Sandler 2007).

5.3 *Final Cases*

Finally, consider the following intelligence representation:

$$\max_{x_i} U_i \left[w_i - x_i c_i \left(x_j \right), I \left(x_i, x_j \right) \right], \tag{24}$$

where i's unit price of intelligence is $c_i(x_j)$ with $dc_i/dx_j > 0$ indicating congestion. A similar representation holds for country j. Intelligence is no longer exclusive so that each country gain not only from its intelligence but also from that of its counterpart—i.e., $I_i > 0$, $I_j > 0$, and $I_{ij} < 0$. With standard procedures, this can be shown to be a case of strategic substitutes with negatively sloped best-response paths. Unlike the previous model, this case can be consistent with plain substitutes or complements depending on marginal crowding cost $\left(-x_i \frac{\partial U_i}{\partial y_i} \frac{dc_i}{dx_j}\right)$ relative to i's marginal gain from j's intelligence effort $\left(\frac{\partial U_i}{\partial I} I_j\right)$. If, say, marginal crowding cost is the stronger influence, $\partial U_i/\partial x_j < 0$, indicative of plain substitutes. This then means that Fig. 4 still applies; otherwise, Fig. 1 is relevant with relabeling of the *IC* curves. In either case, leadership and/or unilateral action is not helpful to both countries.

There are other scenarios. For example, the following formulation,

$$\max_{x_i} U_i \left[w_i - c_i x_i, x_i, I\left(x_i + x_j\right)\right], \tag{25}$$

where c_i is the price parameter. This formulation permits positive private and public joint products to be derived from *i*'s intelligence. For intelligence, the aggregator technology can also come into play (Cornes 1993). Obtaining intelligence on a terrorist group is a best-shot public good, dependent on the greatest effort, while maintaining the secrecy of gathered intelligence is a weakest-link or weaker-link public good, more dependent on the smaller concealment efforts. This aggregator technology affects the form of the *I* function and presence of strategic or plain substitutes and/or complements.

6 Terrorists Versus Government

I conclude with an analysis of adversaries—a targeted government (g) and the terrorist group (t)—for which there can be a rich set of counterterrorism externalities (Gaibulloev et al. 2017). The model is stripped down to its bare essentials. However, even in its primitive form, there is a hybrid case of strategic substitutes and strategic complements.

The terrorist group's problem is to maximize utility (u) minus costs (c):

$$\max_a \left[u\left(a, \varphi\right) - c\left(a, e\right)\right], \tag{26}$$

in which a denotes terrorist attacks, e represents the government's counterterrorism efforts, and φ is the group's radicalization parameter. Terrorists' utility derives from their induced casualties and property damage, which increase with attacks, so that $u_a > 0$ and $u_{aa} < 0$. Increased radicalization augments the marginal utility of attacking, so that $u_{a\varphi} > 0$. Terrorists' cost satisfies the following: c_a, c_e, c_{aa}, c_{ee}, and $c_{ea} > 0$, for which the first four partials indicate that cost rises at an increasing rate in terms of increased attacks or enhanced counterterrorism measures. The cost cross-partial is positive because greater counterterrorism measures lift the marginal cost of attacks. With this set of assumptions, the slope of the terrorist best-response path, BR_t, is

$$\frac{\partial BR_t}{\partial e} = \frac{c_{ae}}{u_{aa} - c_{aa}} < 0, \tag{27}$$

indicative of strategic substitutes. The denominator of (27) is the SOC, which is negative. This can be shown to be a situation of plain substitutes because $\partial (u - c) / \partial e = -c_e < 0$, based on the objective associated with (26). Hence, the terrorists' isoprofit curve, IPC_t, is an inverted C-shape. Unilateral efforts at increased radicalization implies

$$\frac{\partial BR_t}{\partial \varphi} = \frac{-u_{a\varphi}}{u_{aa} - c_{aa}} > 0, \tag{28}$$

or an upward shift in the terrorists' best-response path.

The adversarial government's problem is

$$\min_{e} \left[\psi l \left(e, a\right) + C(e)\right], \tag{29}$$

or to choose its counterterrorism to minimize the sum of attack-induced loss, l, plus counterterrorism cost, C. The government puts a ψ weight on losses, so that more democratic governments are anticipated to weigh such losses greater. The loss function abides by the following reasonable assumptions: $l_e < 0$, $l_{ee} > 0$, $l_a > 0$, $l_{aa} > 0$, and $l_{ea} < 0$. The cross-partial indicates that the marginal loss from an attack is ameliorated by counterterrorism measures that may involve enhanced first-responder capabilities.

The slope of the government's best-response curve, BR_g, is

$$\frac{\partial BR_g}{\partial a} = \frac{-\psi l_{ea}}{\psi l_{ee} + C''} > 0, \tag{30}$$

where the denominator is positive owing to the satisfaction of a minimum's SOC. Thus, the government's best-response function is positively sloped indicative of strategic complements. Differentiation of the objective in (29) with respect to attacks is consistent with plain substitutes, so that the government's isocost contours, IC_g, are inverted U-shaped curves.

This mixed case of strategic substitutes and complements possesses interesting implications for leadership and unilateral action that differ from all previous cases. The Nash equilibrium, E, is at the intersection of BR_g and BR_t in Fig. 5, where the shaded lens-shaped region consists of Pareto-superior points to E. Leadership by the government result in S_g, for which the government's well-being is improved at the expense of the terrorist group. However, leadership by the terrorist group results in S_t, for which the well-being of both adversaries improves. In Fig. 5, terrorist leadership ratchets down the hostilities, leading to fewer attacks and less counterterrorism measures, which improve both adversaries' well-being. This scenario occurs when terrorists lead by limiting attacks—e.g., Fuerzas Armadas Revolucionarias de Colombia (FARC) in recent times. Mixed cases imply asymmetric leadership outcomes.

Unilateral action by the terrorist group is harmful to the government if it involves increased radicalization, which shifts BR_t in the northeast direction. This

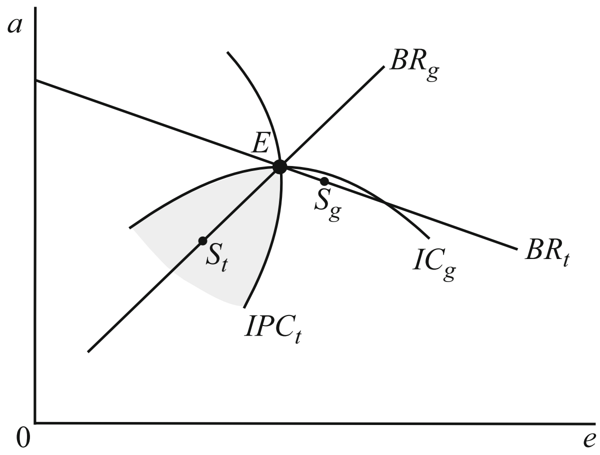


Fig. 5 Mixed case: terrorist versus government

augmented radicalization changes the position of all of the terrorists' IPC_t curves due to a change in the group's utility function; hence, comparison to the IPC_t through E in Fig. 5 is not informative. If, however, unilateral action corresponds to decreased radicalization owing to new leadership or government concessions, then the government's well-being improves.

For the government, unilateral action corresponds to changes in ψ or the value it places on citizen's safety. The effect of this parameter on BR_g is

$$\frac{\partial BR_g}{\partial \psi} = \frac{-l_e}{\psi l_{ee} + C''} > 0, \tag{31}$$

or an eastward shift in BR_g in Fig. 5, which improves the government's well-being at the expense of the terrorists. The mixed strategic scenario allows leadership and unilateral action to imply different outcomes. Unlike the case examined here, there may exist terrorist groups that augment their attacks in response to greater government counterterrorism, so that BR_t is positively sloped. This may correspond to risk-loving groups. Throughout the chapter, terrorists are assumed to be risk neutral, but this implicit assumption could be altered.

Thus far, only one stage to the game is permitted. Multiple stages prove useful when the counterterrorism game is extended to more agents, such as voters or additional targeted governments (Sandler and Siqueira 2009; Siqueira and Sandler 2007). For example, the recruitment of terrorist operative may constitute the first stage; the interaction between the terrorist group and the government characterizes the second stage; and the voter's reaction to government success or failure involves the third stage.

7 Concluding Remarks

Cornes and his co-authors greatly enriched the theoretical foundations of the private provision of public goods and the analysis of externalities. In a modest way, this chapter highlights the practical importance of these contributions by applying a tiny portion of them to the study of terrorism and counterterrorism. The variety of applications underscores the richness of this work. These principles can be applied to the study of regional and global public goods, including the study of transfrontier pollution (Peinhardt and Sandler 2015; Sandler 2004). As such they have much to say about climate change, ozone-shield depletion, acid rain, and deforestation. The work of Cornes and associates also enlightens us about global health, world security, global governance, and knowledge generation.

References

Arce, D. G., & Sandler, T. (2007). Terrorist signaling and the value of intelligence. *British Journal of Political Science, 37*(4), 573–586.

Bandyopadhyay, S., & Sandler, T. (2011). The interplay between preemptive and defensive counterterrorism measures: A two-stage game. *Economica, 78*(3), 546–564.

Bandyopadhyay, S., & Younas, J. (2011). Poverty, political freedom, and the roots of terrorism in developing countries: An empirical assessment. *Economics Letters, 112*(2), 171–176.

Bergstrom, T. C., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics, 29*(1), 25–49.

Buchholz, W., & Sandler, T. (2017, forthcoming). Successful leadership in global public good provision: Incorporating behavioral approaches. *Environmental and Resource Economics.*

Bulow, J. L., Geanakoplos, J. D., & Klemperer, P. D. (1985). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy, 93*(3), 488–511.

Cárceles-Poveda, E., & Tauman, Y. (2011). A strategic analysis of the war against transnational terrorism. *Games and Economic Behavior, 71*(1), 49–65.

Cornes, R. (1993). Dyke maintenance and other stories: Some neglected types of public goods. *Quarterly Journal of Economics, 108*(1), 259–271.

Cornes, R., & Hartley, R. (2007). Aggregative public good games. *Journal of Public Economic Theory, 9*(2), 201–219.

Cornes, R., Hartley, R., & Sandler, T. (1999). Equilibrium, existence and uniqueness in public goods models: An elementary proof via contraction. *Journal of Public Economic Theory, 1*(4), 499–509.

Cornes, R., & Sandler, T. (1983). On commons and tragedies. *American Economic Review, 73*(4), 787–792.

Cornes, R., & Sandler, T. (1984). Easy riders, joint production, and public goods. *Economic Journal, 94*(3), 580–598.

Cornes, R., & Sandler, T. (1985). The simple analytics of pure public good provision. *Economica, 52*(1), 103–116.

Cornes, R., & Sandler, T. (1986). *The theory of externalities, public goods, and club goods*. Cambridge: Cambridge University Press.

Cornes, R., & Sandler, T. (1994). The comparative static properties of the impure public good model. *Journal of Public Economics, 54*(3), 403–421.

Cornes, R., & Sandler, T. (1996). *The theory of externalities, public goods, and club goods* (2nd ed.). Cambridge: Cambridge University Press.

Dreher, A., Gaston, N., & Martens, P. (2008). *Measuring globalization—Gauging its consequence*. New York: Springer.

Eaton, B. C. (2004). The elementary economics of social dilemmas. *Canadian Journal of Economics, 37*(4), 805–829.

Eaton, B. C., & Eswaran, M. (2002). Asymmetric equilibria and strategic behavior in oligopolistic markets. In B. C. Eaton (Ed.), *Applied microeconomic theory*. Cheltenham, UK: Edward Elgar.

Enders, W., Hoover, G. A., & Sandler, T. (2016). The changing nonlinear relationships between income and terrorism. *Journal of Conflict Resolution, 60*(2), 195–225.

Enders, W., & Sandler, T. (2012). *The political economy of terrorism* (2nd ed.). Cambridge: Cambridge University Press.

Enders, W., Sandler, T., & Gaibulloev, K. (2011). Domestic versus transnational terrorism: Data, decomposition, and dynamics. *Journal of Peace Research, 48*(3), 319–337.

Eubank, W. L., & Weinberg, L. B. (1994). Does democracy encourage terrorism? *Terrorism and Political Violence, 13*(1), 155–164.

Eyerman, J. (1998). Terrorism and democratic states: Soft targets or accessible systems? *International Interactions, 24*(2), 151–170.

Gaibulloev, K., Piazza, J. A., & Sandler, T. (2017, forthcoming). *Regime types and terrorism. International Organization*, Richardson, TX.

Gardeazabal, J., & Sandler, T. (2015). INTERPOL's surveillance network in curbing transnational terrorism. *Journal of Policy Analysis and Management, 34*(4), 761–780.

Gassebner, M., & Luechinger, S. (2011). Lock, stock, and barrel: A comprehensive assessment of the determinants of terror. *Public Choice, 149*(3–4), 235–261.

Hoel, M. (1991). Global environmental problems: The effects of unilateral actions taken by one country. *Journal of Environmental Economics and Management, 20*(1), 55–70.

Hoffman, B. (2006). *Inside terrorism* (revised ed.). New York: Columbia University Press.

Krueger, A. B., & Maleckova, J. (2003). Education, poverty, and terrorism: Is there a causal connection? *Journal of Economic Perspectives, 17*(4), 119–144.

Li, Q., & Schaub, D. (2004). Economic globalization and transnational terrorism. *Journal of Conflict Resolution, 48*(2), 230–258.

Peinhardt, C., & Sandler, T. (2015). *Transnational cooperation: An issue-based approach*. New York: Oxford University Press.

Piazza, J. A. (2007). Draining the swamp: Democracy promotion, state failure, and terrorism in 19 Middle Eastern countries. *Studies in Conflict and Terrorism, 30*(6), 521–539.

Piazza, J. A. (2008). Do democracy and free markets protect us from terrorism? *International Politics, 45*(1), 72–91.

Rosendorff, B. P., & Sandler, T. (2004). Too much of a good thing? The proactive response dilemma. *Journal of Conflict Resolution, 48*(5), 657–671.

Sandler, T. (1995). On the relationship between democracy and terrorism. *Terrorism and Political Violence, 7*(4), 1–9.

Sandler, T. (2004). *Global collective action*. Cambridge: Cambridge University Press.

Sandler, T. (2014). The analytical study of terrorism: Taking stock. *Journal of Peace Research, 51*(2), 257–271.

Sandler, T., & Lapan, H. E. (1988). The calculus of dissent: An analysis of terrorists' choice of targets. *Synthese, 76*(2), 245–261.

Sandler, T., & Siqueira, K. (2006). Global terrorism: Deterrence versus preemption. *Canadian Journal of Economics, 50*(4), 1370–1387.

Sandler, T., & Siqueira, K. (2009). Games and terrorism: Recent developments. *Simulation & Gaming, 40*(2), 164–192.

Schneider, F., Brück, T., & Meierrieks, D. (2015). The economic survey of counterterrorism: A survey. *Journal of Economic Survey, 29*(1), 131–157.

Siqueira, K. (2005). Political and militant wings within dissident movements and organizations. *Journal of Conflict Resolution, 49*(2), 218–236.

Siqueira, K., & Sandler, T. (2006). Terrorists versus the government: Strategic interaction, support, and sponsorship. *Journal of Conflict Resolution, 50*(6), 878–898.

Siqueira, K., & Sandler, T. (2007). Terrorist backlash, terrorism mitigation, and policy delegation. *Journal of Public Economics, 91*(9), 1800–1815.

Linguistic Assimilation and Ethno-religious Conflict

Indraneel Dasgupta

1 Introduction

In most, arguably all, diverse societies with a dominant majority ethno-linguistic community, public policy debates reflect a recurring contestation between mono-culturalism and multi-culturalism. In the first case, minorities are supposed to 'assimilate', i.e., adopt the cultural-linguistic conventions and behavioural patterns of the majority (at least over time). In the second case, minorities are to be permitted, perhaps even encouraged, to articulate and develop their distinct cultural-linguistic identities. This paper provides an analytical framework within which these policy stances can be assessed, and their implications for *ethno-religious* social conflict and income distribution explicated.

Policies to incentivise individual members of minority communities to embrace majority ethno-linguistic conventions are pervasive across the world. Language, syllabus and cultural policies followed in public educational institutions, the official language followed in law courts and public administration, language and cultural content of citizenship tests, etc., are all instruments that can and indeed are used to

Many of the ideas explored in this paper have their roots in conversations with the late Richard Cornes, but no acknowledgement is adequate for my intellectual debt to him, accumulated over many years. I have also benefited immensely from detailed comments by an anonymous referee on a previous draft.

I. Dasgupta (✉)
Economic Research Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata, 700108 West Bengal, India
e-mail: indraneel@isical.ac.in

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_12

nudge minority individuals towards extensive adoption of majority ethno-linguistic conventions, by increasing the relative benefits from doing so.[1]

Arguments for linguistic assimilation often start from the presumption that minorities lead a segregated existence, whether literally or in terms of social interactions. Such minority communities may have had a long historical presence (as is the case with indigenous communities in countries of European settlement), or may be the product of recent immigration (as for example is the case in Western Europe). Frankly articulated presumptions of divine sanction or inherent cultural superiority aside, the broad contours of an instrumentalist case for assimilation appear to be the following.

First, linguistic assimilation is profitable for minorities, since adoption of the cultural and linguistic practices of the majority reduces search, coordination and transaction costs contingent on economic interaction with the latter. This in turn expands the effective size of the market, generating efficiency gains via specialization, economies of scale and faster adoption of social, institutional and technological innovations. For the same reasons, assimilation by the minority is also profitable for the majority. Second, cultural/linguistic diversity encourages and empowers incompatible belief systems or historical identifications (i.e., it buttresses oppositional *ethno-religious* identities). The latter both perpetuate atavistic antagonisms and generate new ‘culture wars’ between communities: assimilation reduces the scope and intensity of such conflicts. Third, cultural-cum-linguistic segregation, by leading to socio-economic exclusion, generates a poverty-stricken minority underclass, which puts pressure on the welfare system and/or law enforcement, thereby negatively impacting the majority. Xenophobic political parties, in particular, often seek to magnify and exploit majority anxieties by simultaneously charging minorities’ with both an unwillingness to assimilate and an excessive propensity to engage in crime, and explain away their poverty and exclusion in such terms.[2]

Despite its policy importance, comparative assessment of the impacts of assimilation and segregation, on income distribution and ethno-religious (or racial) conflict, has received little analytical attention in the formal theoretical literature on

[1]Denial of recognition to the Kurdish language in Turkey is linked to the Turkish nationalist policy of cultural assimilation. In Latvia, despite about 40 % of the population being Russian-speaking, Latvian remains both the sole state language and a requirement for citizenship. In the UK, English language requirements for citizenship tests have been progressively tightened in recent years. Ortega and Tangerås (2008) develop a political-economic analysis of the imposition of monolingual education by dominant groups.

[2]The first extends a free trade argument into social policy (e.g., Lazear 1999). It was advanced by colonial administrators and social reformers in colonized countries in the nineteenth and early twentieth centuries, as the justification for westernizing the education system, the legal code, and social behaviour. Civil rights laws and anti-discrimination statutes in the US are motivated at least partly by the belief that social integration promotes economic efficiency (Fredrickson 1999). Contemporary examples of populist articulation of the second and third arguments include political parties such as the French National Front, the Dutch Party for Freedom, the Bharatiya Janata Party of India, Jobbik of Hungary and Golden Dawn of Greece.

political economics.[3] This paper seeks to address this lacuna. I consider a society consisting of a majority and a (relatively large) minority. These communities differ in terms of their characteristics, acquired as part of childhood socialization of community members, along two different dimensions. One set of characteristics are directly relevant for economic interaction and productivity, while the other set involves intrinsic valuation of certain items, practices, or symbols that do not have any direct productivity implications. To fix ideas, one may concretize this dichotomy broadly as that between *language* on the one hand and *religion or race* on the other.[4] The first set of characteristics is, in principle, open to change on the basis of individual adjustments to economic incentives. Assimilation, in my model, therefore takes the form of individually rational minority adoption of majority practices in this sphere, which in turn has consequences for income distribution. The second set of characteristics is however more deeply or foundationally constitutive of one's sense of self, and therefore stable. This constitutes the site of collective conflict between communities in my model. The degree of assimilation affects such conflict via its determination of income distribution.

To expand, individuals acquire a set of cultural-behavioural traits and conventions, as part of their upbringing within a particular community, which are relevant for workplace interaction and coordination. One achieves income gains when a larger proportion of the workforce comes to share one's behavioural patterns and expressive conventions, thereby facilitating economic coordination (e.g., Schelling 1960; Lewis 1969). Language, including dialect, idiom, accent and modes of expression, constitutes the most transparent example of such productivity-relevant conventions. Expanding on Akerlof and Kranton (2000), I assume that 'switching identity', or bringing one's behaviour into alignment with those commonly present in (and thereby constitutive of) the other community, is feasible but costly. Individuals vary in terms of their identity switching costs. Thus, for a minority individual,

[3]Lazear (1999), Kónya (2005), Kuran and Sandholm (2008), Li (2013) and Bowles et al. (2014) develop models of assimilation (or, more generally, social segregation and integration), but do not analyze the implications for ethnic conflict. Akerlof and Kranton (2000) explain forms of dysfunctional individual behaviour in terms of stresses generated by identity norms, but do not model their aggregate consequences for conflict between communities. Conversely, Dasgupta and Kanbur (2005b, 2007) and Esteban and Ray (2008, 2011) examine how *exogenous* changes in the income distribution affect conflict between communities, and thus do not connect the income distribution to the extent of cultural-linguistic integration prevailing in the society. The connection between cultural integration and social conflict thus remains unexplored in their analysis. Dasgupta (2009) shows how class conflict between workers and employers, and ethnic conflict between different groups of workers, mutually condition one another, but assumes homogeneity within the working class in all employment-relevant aspects except the reservation wage. Bakshi and Dasgupta (2017) explore the dynamic evolution of ethnic conflict when state institutions are susceptible to ethnic capture. Our analysis has a distant family resemblance with that of Bhattacharya et al. (2015), who examine the role of inter-group mobility in the emergence of conflict, but varies greatly from theirs both in its specific structure and modelling of conflict.

[4]In reality, some practices with religious, racial or ethnic identity connotations may also have productivity implications. The broad-brush distinction is porous but nonetheless empirically helpful, and is routinely deployed in the economic analysis of discrimination.

the decision whether to assimilate (i.e., to exhibit the majority's workplace-relevant behavioural traits and conventions) is guided by the relative return from doing so, net of her identity switching cost. This net relative return is, in turn, determined by the proportion of her community members who choose to assimilate. I first show that, when the minority is relatively large, assimilation by the entire minority community, and separation (i.e., a complete lack thereof) both constitute locally stable equilibria. Thus, one-off 'big push' policies, which force a large proportion of minority individuals to integrate, can permanently convert a culturally-linguistically segregated society into one integrated within the sphere of production-relevant behavioural and expressive conventions. Such a move increases both total income and that of every majority individual; total income of the minority community may also rise. However, assimilation worsens the income distribution both across and within communities: it reduces the income share of the minority community, while also increasing income inequality therein.

I proceed to explicate the consequences for collective conflict. Building on Dasgupta and Kanbur (2005a, 2007, 2011), I visualize communities as held together by certain forms of community-specific public goods, such as institutions and rituals of public collective worship, historical monuments, public statues and memorials to military leaders, political icons and past victories, laws governing behaviour in private matters of sexuality, marriage, divorce, inheritance, abortion, etc. These carry no relevance for productivity, and are therefore orthogonal to workplace-relevant (largely cultural-linguistic) conventions, but are intrinsically valued (mostly on core ethno-religious or racial identitarian grounds). In accord with Esteban and Ray (2008, 2011), I model such collective consumption as generating political conflict between communities over mutually exclusive control of the public sphere.[5] All members of a community derive non-material benefits from, and may therefore contribute material political resources to, making the public sphere more reflective of its collective symbols and values. Thus, conflict-inducing political expenditure takes on the characteristic of a group-specific pure public good, generated by

[5] In 2010, France banned the wearing of a face-covering veil in public. The key official justification was productivity-relevant: face-coverings prevent identification, which is both a security risk and a coordination hindrance, in a society which relies on facial recognition and expression in communication. Thus, headscarves were not affected. In contrast, the wearing of *all* conspicuous religious symbols in public schools was banned in France in 2004 by a different law, which did affect the wearing of both Islamic veils *and* headscarves. The Turkish government has traditionally banned women who wear headscarves from working in the public sector. In both cases, the ban on headscarves was justified not by any direct negative impact it might have on productivity, but by its symbolic role in keeping the public sphere secular. In a referendum held in Switzerland in 2009, a constitutional amendment banning new mosque minarets was approved by 57.5 %. In Northern Ireland, clashes often break out over rival Catholic and Protestant marches organized annually to commemorate events in the history of past antagonisms, in India Hindus and Muslims contest ownership of medieval structures, while in Europe conflicts rage between rival mobilizations over demands for censorship on grounds of blasphemy. In all these cases, the items of contestation do not appear to have any direct or immanent implications for workplace coordination or economic productivity, but are intrinsically valued by (typically ethno-religious) communities as constitutive symbols of self-expression in the public sphere.

decentralized voluntary contributions. This feature of the model builds on the canonical literature on voluntary contributions to pure public goods, stemming from the seminal work of Bergstrom et al. (1986) and Cornes and Schweinberger (1996). I show that the inequality inducing consequences of assimilation in the workplace *spill over* from the sphere of private incomes to that of public assertion: assimilation reduces the minority's share of the symbolic and normative content of the public sphere, thus reducing the welfare of at least some minority individuals. Indeed, all minority individuals may be worse off in consequence. However, assimilation in the workplace may reduce *relative* social waste due to political conflict, measured as the share of social income expended on political activities, even while possibly increasing it in absolute terms.

Section 2 sets up the benchmark model. I examine conflict over collective consumption in Sect. 3. Section 4 concludes with a discussion of some extensions and applications. Detailed proofs are presented in the Appendix.

2 The Model

2.1 Preliminaries

Consider a population of size normalized to 1, comprised of two groups, M (majority) and N (minority), with population shares m and n respectively, $m = (1-n)$, $n \in \left(0, \frac{1}{2}\right)$. Each member of the population is endowed with one unit of effort, which she expends on activities related to earning income. Subsequent to earning income, she expends her income on consumption goods, so as to maximize her utility. Consumption decisions have no implications for earning ability in our model, and, ceteris paribus, individual utility will be increasing in own income. Hence we can analyse the prior income earning process independently of subsequent consumption decisions.

To earn income, each individual needs to acquire some identity-related, or community-specific, cultural characteristics including linguistic ones, to successfully engage in production-related transactions, negotiations and coordination. The marginal product of effort, contingent on acquiring the characteristics specific to community $i \in \{M, N\}$ and choosing to exhibit them, is $\theta_i y_i$, where $\theta_i \in [0, 1]$ is the proportion of the population that behaves according to the work-place relevant cultural-cum-behavioural conventions and characteristics of community i, and $y_i > 0$ is some community-wide productivity parameter. Identities are exclusive, or 'oppositional': the exhibition of characteristics of community i implies renunciation of the identity-markers of the other community. Thus, the benefit from displaying a particular set of behavioural patterns depends positively on how pervasive those behavioural patterns are. This captures the idea that common behavioural and expressive conventions coordinate productive activities across individuals and thereby increase output, as in Schelling (1960) and Lewis (1969). These conventions may possibly have intrinsic consequences for productivity as well: y_M does not need to be equal to y_N. I normalize y_M to unity, and assume $y_N \leq 1$. Given any community $i \in \{M, N\}$, I shall denote the other community by $-i$.

For j born into community i, acquisition of the behavioural and expressive conventions of her own community is costless (reflecting socialization in childhood), but acquisition of those of the other community involves an 'identity switching' cost, modelled as an effort cost c; c is idiosyncratic and distributed over $[\rho_i, \overline{\rho}_i]$, with $0 < \rho_i < \overline{\rho}_i < 1$, according to some continuous and differentiable distribution function $F^i(c)$.

An obvious interpretation of c is in terms of the effort spent in learning a new language and behavioural rules instead of engaging in actual production: some are inherently more efficient learners. A deeper one is that not all can internalize alien conventions equally. The degree of functionality within the context of a set of culturally/linguistically alien rules varies across persons born into the same community, leading to idiosyncratic differences in productivity. These differences are however not intrinsic but specific to the cultural construction of the workplace: these differences would disappear if production was organized according to the conventions one was originally socialized into. In any case, the formal upshot is that, for j born into community $-i$, the return from adopting the production-relevant behavioural conventions of the other community, i, is $\theta_i \left(1 - c_{-i,j}\right) y_i$, where $c_{-i,j}$ is the identity-switching (marginal) effort cost of working in an alien environment for the individual.[6] For such an individual, the return from persisting with one's original behavioural conventions is $(1 - \theta_i)\, y_{-i}$. I assume that the distribution of identity switching costs follows a strictly concave exponential form:

$$F^i(c) = (\overline{\rho}_i - \rho_i)^{-\alpha_i}(c - \rho_i)^{\alpha_i}; \tag{1}$$

where $\alpha_i \in (0, 1)\ \forall i \in \{M, N\}$. The concavity assumption implies that more than half the minority population falls below the mid-point of the cost distribution. Thus, intuitively, minority individuals are more likely to be low cost, rather than high cost; or, equivalently, concentrated in the lower part of the cost distribution with regard to assimilation. The opposite holds for a convex cost distribution. Thus, a concave cost distribution would appear, a priori, to be the case where assimilation is most likely to benefit the minority community on average. This is why I focus on this case.

[6] Generalized discrimination against the minority can be modelled as a constant cost component, $d \leq \rho_N$, that impacts all assimilating N individuals equally. Thus, an increase in such discrimination simply reduces the returns from assimilation by an identical amount for all minority individuals. Individuals may perceive their own expressive and behavioural habits as norms rather than conventions, in that they may intrinsically value them as ideals to live by. In that case, identity-switching will involve a psychic cost. If such marginal psychic cost increases with the level of workplace effort, individuals may rationally provide less than full effort in an alien work environment. The effort level provided will then vary according to idiosyncratic differences in the marginal psychic cost function. Though evidently compatible with my analysis, I refrain from explicitly modelling this additional source of idiosyncratic differences in productivity on considerations of expositional ease and simplicity.

However, as I shall discuss later (in the context of Proposition 1), my substantive results will hold regardless of the concavity assumption.

Contingent on switching identity, the income $I_{-i,j}$ of j born into community $-i$ falls in the interval $[\theta_i (1 - \overline{\rho}_{-i}) y_i, \theta_i (1 - \rho_{-i}) y_i]$; when the entire community $-i$ switches identity, the distribution of (normalized) income within that community is given by:

$$D^{-i}\left(\frac{I}{\theta_i y_i}\right) = 1 - F^{-i}\left(1 - \frac{I}{\theta_i y_i}\right).$$

Let n_M be the size of the 'assimilated' minority population (those who choose to exhibit the behavioural and expressive conventions of the majority despite being brought up in the minority community); $n_M \in [0, n]$. Then the assimilation cost of the marginal assimilated member of N is given by:

$$\breve{c}\,(n_M) \equiv F^{N^{-1}}\left(\frac{n_M}{n}\right). \tag{2}$$

$\breve{c}(.)$ is the inverse supply function for assimilated individuals: if the population size of N individuals who rationally assimilate is n_M, then the highest cost incurred must be exactly $\breve{c}\,(n_M)$. By (1) and (2):

$$\breve{c}\,(n_M) \equiv \left(\frac{n_M}{n}\right)^{1/\alpha_N} (\overline{\rho}_N - \rho_N) + \rho_N; \tag{3}$$

so that (recalling $\alpha_N \in (0, 1)$):

$$\breve{c}'\,(n_M) = \frac{(\overline{\rho}_N - \rho_N)}{n\alpha_N}\left(\frac{n_M}{n}\right)^{\frac{1-\alpha_N}{\alpha_N}} > 0 \text{ for all } n_M \in (0, n]\,; \tag{4}$$

$$\breve{c}''\,(n_M) = (1 - \alpha_N)\,\frac{(\overline{\rho}_N - \rho_N)}{(n\alpha_N)^2}\left(\frac{n_M}{n}\right)^{\frac{1-2\alpha_N}{\alpha_N}} > 0. \tag{5}$$

Thus, the marginal assimilation cost function (or the inverse supply function) $\breve{c}(.)$ is *increasing and convex* in the size of the assimilated population over $(0, n]$. Analogous expressions hold for M.

2.2 Equilibrium

Individuals simultaneously decide whether to acquire the behavioural traits of the other community or to persist with their own, i.e., those they are already endowed with, with the objective of maximizing their own income. A Nash equilibrium of this income maximizing game is simply a set of identity choices such that the choice made by any individual maximizes her own income, given those of all other individuals in society.

Since $\rho_i > 0$ for all $i \in \{M, N\}$, if at least one member of community i earns at least as much by switching, then all members of the other community $(-i)$ must earn more by continuing with their own identity. Thus, apart from the two possible monocultural outcomes, where all individuals choose to exhibit identical behavioural traits and conventions, I only need to consider the class of 'multicultural' outcomes where all members of some community i maintain their own communal identity markers, and at least some (possibly all) members of the other community –*i* persist with the identity markers of that community (-*i*), as possible candidates for Nash equilibrium.

I shall first consider assimilation by minority (N) individuals to majority (M) norms. In light of the preceding discussion, any given level of assimilation $n_M \in [0, n]$ constitutes an equilibrium if, given that level of assimilation and persistence of all M individuals with their own cultural traits, (a) n_M/n proportion of N individuals are all at least as well off by assimilating, and (b) the remaining proportion of N individuals are all at least as well off by *not* assimilating. Note that condition (a) above implies that all M individuals are better off by persisting with their own behavioural traits when $n_M > 0$. When $n_M = 0$, given the persistence of all other M individuals with their own cultural traits, every M individual is worse off in case of a unilateral deviation to the minority's traits (since $m > n$). An equilibrium n_M^* is (locally) stable if there exists $\varepsilon > 0$ such that: [for all $n_M \in \left(n_M^*, n_M^* + \epsilon\right)$, more than $\left(1 - \frac{n_M}{n}\right)$ proportion of N individuals are worse off by assimilating; and, for all $n_M \in \left(n_M^* - \epsilon, n_M^*\right)$, more than $\frac{n_M}{n}$ proportion of N individuals are better off by assimilating]. Evidently, since $n_M \in [0, n]$, the first part of the stability condition above must hold vacuously when $n_M^* = n$, while the second part must hold vacuously when $n_M^* = 0$. An equilibrium where M individuals acquire the minority's behavioural norms is defined analogously.

I now impose a restriction via Assumption 1 below. Assumption 1 formalizes the intuitive idea that assimilation costs are substantial *relative* to the size of the majority, and thus prohibitive relative to private gains from *unilateral* assimilation on part of N individuals. This does not prevent assimilation costs from being arbitrarily small for a positive proportion of the N population: while positive, ρ_N can be arbitrarily close to 0. However, the closer ρ_N to 0, the closer the majority must be in size to the minority.

Assumption 1 $\left[m < \frac{y_N}{1-\rho_N+y_N}\right]$.

Proposition 1 *Let Assumption 1 hold. Then, exactly three locally stable equilibria exist, two of which entail behavioural uniformity, while one entails complete behavioural separation.*

Proof See the Appendix.

By Proposition 1, only three stable equilibria exist. One involves assimilation of the entire minority community to the majority's conventions: the latter thus universally prevail in this mono-cultural equilibrium. However, a stable multi-cultural equilibrium also exists, where *all* persist with the cultural-linguistic conventions specific to their own respective communities. Lastly, assimilation of the entire majority to the minority's conventions constitutes a locally stable mono-cultural equilibrium as well.[7]

One can neither rule out, nor ensure, a stable equilibrium involving assimilation by a part, but not all, of the minority community if the cost distribution is convex (i.e., if $\alpha_N \geq 1$). However, even in these cases, the three cases discussed under Proposition 1 all continue to constitute locally stable equilibria, given Assumption 1. Thus, complete assimilation by the minority and complete behavioural separation between the two communities constitute locally stable equilibria regardless of any restriction on the values of α_N and α_M. Since my subsequent results presented in Propositions 2 and 3 below only involve comparisons of the properties of these two equilibria, it follows that they also hold irrespective of the values of α_N and and α_M.

Proposition 1 implies that the same two communities can get locked into either a mono-cultural or a multi-cultural equilibrium, depending on accidents of past history. Thus, if a minority currently exists in a state more or less culturally separate from the majority, that cannot, by itself, be construed as evidence of that community's inherent or constitutive inability to assimilate. Rather, it can be perceived as the result of a collective action problem: a coordinated attempt at assimilation, if sufficiently widespread within the minority community, may indeed succeed in completely assimilating that community to majority conventions. Conversely, an assimilated minority may successfully construct its separate identity through a coordinated attempt at cultural assertion and 'invention of traditions'.[8] In either case, Proposition 1 suggests that assimilation or separation may be a collective choice in a broad sense, rather than a social given. The state, through its language, schooling, employment and citizenship policies, may be able to exercise that choice effectively. Once exercised on a sufficiently large scale, the outcome would be self-sustaining. Since the society would shift to a different locally stable equilibrium, compulsion would no longer be necessary: voluntary decentralized individual choice would continue to reproduce the desired outcome. A similar big push may be exercised by large-scale social movements inside the minority community as well.

Since Proposition 1 suggests that separation or assimilation may be a matter of collective choice, it also highlights the importance of comparing the characteristics of these alternate equilibria in the formulation of social policy. I now proceed to address this question.

[7]There exist multi-cultural equilibria involving partial assimilation as well, but these are all unstable.

[8]For a detailed discussion of such movements and their role in the construction of ethno-linguistic nationalism in modern Europe, see Hobsbawm and Ranger (1983) and Hobsbawm (1992).

2.3 Comparing Equilibria

I compare the properties of the equilibrium where the minority assimilates, with those of the equilibrium where all persist with their original cultural-linguistic conventions, so that the communities fully maintain their cultural cum behavioural and linguistic 'separateness'. For brevity, I shall term the first, 'assimilation' and the second, 'separation'. For the rest of this paper, I ignore the locally stable equilibrium where the majority assimilates to minority conventions since this appears generally devoid of substantive policy interest: it is difficult to think of societies where such an equilibrium may be thought to obtain.[9]

Proposition 2 *Let Assumption 1 hold. Then, under assimilation by the minority community relative to separation:*

(a) *every member of the majority community earns more, and the majority community's share of total income rises;*
(b) *total income in society is higher;*
(c) *total income of the minority community rises iff* $[(1 - ny_N) > E(c_N)]$*; its total income is reduced if the inequality is reversed; and*
(d) *all minority individuals suffer an absolute income reduction if* $[(1 - ny_N) < \rho_N]$ *while* at least *some do so if* $[(1 - ny_N) < \overline{\rho}_N]$.

Proof See the Appendix.

Proposition 2 articulates the efficiency argument for assimilation. Every member of M gains income if N assimilates. The economies of scale generated by assimilation outweigh the costs of integration incurred by the latter, so that total income of society necessarily increases. However, assimilation also leads to increased inequality along two different dimensions. First, it benefits M *proportionately more*. Second, while incomes within a community are identical under separation, reflecting equal inherent productivity, idiosyncratic differences in the ability to function within an alien culture opens up income inequality inside N when it assimilates (though incomes within M remain equalized).

Despite a decline in income share, N benefits monetarily on average from assimilation when the gain from assimilation is greater than the average cost. The larger the majority and the lower the relative productivity of the minority, the higher this gain. However, provided that the upper bound on assimilation costs is higher than the gain from assimilation, a positive proportion of N individuals (those with costs in $(1 - ny_N, \overline{\rho}_N)$) must suffer a fall in income under assimilation. When costs are sufficiently high relative to the gains ($1 - ny_N < \rho_N$), assimilation reduces the income of *every* N individual.

[9] Furthermore, such an equilibrium necessarily generates lower total output than the one where the minority assimilates, provided M, on average, finds it at least as costly to change its behavioural patterns as the minority (i.e., $E(c_M) \geq E(c_N)$], and may do so even otherwise. Since this is empirically likely, it is therefore of limited normative interest.

3 Public Consumption and Collective Conflict

I now proceed to incorporate social conflict in the model and investigate how the inequality-inducing consequences of assimilation impact on it. The form of conflict I address is that over collective assertions of religious (and also possibly racial) identity via symbolic domination of the public sphere. A religious community finds its collective identity in shared religious shrines, monuments to its departed heroes, public memory rituals of past victories and defeats, in the naming of parks, streets, bridges, towns and universities after its revered members, mass public gatherings to perform collective religious rituals, state holidays on occasions important to its perceived collective history, etc. Analogous considerations apply to racial communities. Laws governing private behaviour of individuals, especially in matters of marriage, sexual behaviour, divorce, abortion and inheritance are typically based on a set of core values and norms [in the sense of behavioural ideals, as for example in Akerlof (1980)] identified with particular religious communities. Thus, in the first case, a sense of collective ownership is derived from the physical presence of a community's symbolic markers of territory in the public space, literally interpreted. In the second case, a sense of collective possession is derived from the state's identification with, or support for, a set of norms or behavioural ideals central to the self-perception of a community, articulated through the use of the state's legal and administrative machinery (its coercive powers) to enforce the observance of these norms by private individuals. Analytically, therefore, the two cases can be treated identically for my purposes.

When a society consists of multiple religious (or racial) communities with a strongly defined sense of collective history, defined especially in terms of past antagonisms, marking of collective territory in the public sphere is liable to generate conflicts. These may take the form of attempts by different communities to lobby/bribe authorities to act in their favour, for and against the status quo, or they may consist of direct action. Direct action may be legal and peaceful. Examples of this include mass subscription drives to build places of worship more imposing (and therefore more assertive) than those of another community, and mass protests against cow slaughter, perceived blasphemy or laws banning polygyny and juvenile marriage. Agitations for the removal of certain statues in universities (as in the recent campaign by Black South African students for the removal of statues of Cecil Rhodes at Oxford and Cape Town) and other public places constitute another example. Direct action may also be illegal and violent. For example, it may involve the mobilization of activists' groups or militias to physically destroy places of worship or monuments belonging to other communities. Such groups may also terrorize other communities to force them to desist from observing certain practices (e.g., consumption of beef, pork or alcohol) or rituals (e.g., the routine bombing of Shia processions and Sufi shrines by Salafists in Iraq and Pakistan). In any case, offensive action by one community, if unchecked by countervailing defensive action by the other community, generates psychic gains for members of the former community, and losses for members of the latter. Group conflict of this kind can be conceptualized as fights over *distribution of symbolic territory*, rather than of material resources.

Such conflict engages real resources, but the consequent gains are directly psychic, i.e., non-material, depending on the extent to which particular types of public goods specific to a community are generated. For formal purposes, one can demarcate them as taking place in the sphere of collective symbolic consumption, rather than in that of private material consumption. The zero-sum nature of such consumption is parsimoniously modelled via a framework where an individual cares both about consumption in her private sphere and the *share* of the public sphere 'owned' by her religious or racial community.

To fix ideas, suppose, in the status quo, the communities are linguistically separated. As already noted, since linguistic assimilation is a locally stable equilibrium, the state can push the minority into linguistic assimilation through its language policy. Would religious (or racial) conflict over sharing of the public sphere expand or diminish if it did so? This is the main question I seek to address.

Let utility of any member j of community $i \in \{M, N\}$ be given by $u_j^i = U^i\left(x_j^i, p_i, t\right)$; where x_j^i is j's private consumption, p_i is the extent of 'cultural ownership' of the contestable public sphere, $p_i \in [0, 1]$, and $t \in (0, 1]$ is a parameter reflecting the proportion of the public sphere open to contestation by religious (or racial) communities. Greater reflection of the *other* community's symbols in the public sphere (lower p_i) reduces the well-being of all members of community i. Intuitively, the proportion $(1 - t)$ of the public sphere is a *secular* (or race-blind) space: neither community's symbols, nor norms, can be reflected here.[10] For algebraic parsimony, I assume that preferences assume the following form:

$$u_j^i = x_j^i \left(\frac{p_i}{1 - p_i} \right)^{t\aleph_i}; \tag{6}$$

with $\aleph_N, \aleph_M > 0$. The extent of collective ownership is defined through a process of political contestation, which requires the expenditure of monetary resources, generated through decentralized voluntary contributions, on the part of both communities. Formally, community i's ownership share of the

[10] I think of this space as one of constitutional/legal guarantees of secularism or race-blindness. A constitution may ban all overt religious content, practices and symbols from the education system and the public sector; personal laws governing marriage, divorce, inheritance, sexual preference and abortion rights may be based entirely on secular principles and violate the traditional norms of *all* religious communities of any significant size present in that society. Analogously, a legal code may ban the overt display of all racial symbols in public places, or the open discussion of racial identities in the public education system.

contestable public sphere is given by the standard ratio-form contest success function:

$$p_i = \frac{\left(\frac{b_i}{b_{-i}}\right)^{\tau}}{1 + \left(\frac{b_i}{b_{-i}}\right)^{\tau}}; \tag{7}$$

where b_i is the total expenditure by community i in political attempts to influence the normative or symbolic content of public space in its favour, b_{-i} is the total such expenditure by the other community and $\tau \in (0, 1]$ is a parameter. Using (6) and (7),

$$u_j^i = x_j^i \left(\frac{b_i}{b_{-i}}\right)^{\sigma_i t}; \tag{8}$$

where $\sigma_i > 0$. Member j of community i has income I_j^i. All individuals simultaneously allocate their income between private consumption and political contribution so as to maximize utility. Specifically, the problem of member j of community i is given by:

$$\underset{b_i}{Max}\, x_j^i \left(\frac{b_i}{b_{-i}}\right)^{\sigma_i t} \text{subject to : (a) } x_j^i + b_i = I_j^i + b_{-j}^i, \text{ and (b) } b_i \geq b_{-j}^i;$$

where b_{-j}^i denotes the political contribution of all members of community i except j.

Suppose the equilibrium levels of political expenditures are b_M^*, b_N^*. Then, since (recalling (8)) utility only depends on relative political expenditure, a strict Pareto-improvement could be implemented if, somehow, political expenditures of both communities were taxed at some community-neutral rate, and the revenue used to subsidize private consumption of all members of society. A social planner who could credibly pre-commit to the equilibrium division of the public sphere would also be able to enforce a strict Pareto-improvement by eliminating political contributions altogether. Thus, I shall interpret the total amount of political expenditure generated in the Nash equilibrium as a measure of both the total social cost of conflict and its intensity.

The FOCs of the optimization problem of the representative contributing member, j, yield:

$$\forall i \in \{M, N\}, \left[\frac{b_i}{x_j^i} = \sigma_i t\right]. \tag{9}$$

The community-specific parameter σ_i reflects the relative weight on collective consumption vis-a'-vis private consumption. Higher σ_i implies a stronger sense of communal identity, relative to a private notion of self-hood. I accordingly term it *community cohesion*: higher levels of community cohesion increase aggregate political spending by a community. Since equilibrium incomes are identical within M regardless of whether N assimilates, and identical within N in the separated equilibrium, all community members must make identical and positive political contributions in these three cases. Let Y_{NM} be the total income of the minority community if it assimilates, and define:

$$\sigma_N^*(t) \equiv \frac{Y_{NM} - n\,(1 - \overline{\rho}_N)}{t\,(1 - \overline{\rho}_N)}. \tag{10}$$

Recalling (9), it can be checked that, given any $t \in (0, 1]$, when the minority community assimilates, *all* minority individuals must make positive political contributions whenever $\sigma_N > \sigma_N^*(t)$.

Proposition 3 *Given any triple* $< m,\ F^N(c),\ F^M(c) >$ *satisfying Assumption 1, any* $t \in (0, 1]$, *and any* $\sigma_N > \sigma_N^*(t)$,

(a) *there exists* $\epsilon > 0$ *such that political spending as a proportion of total income is lower under assimilation by the minority community, compared to separation, whenever* $\sigma_M < \sigma_N + \epsilon$;
(b) *there exists* $\epsilon > 0$ *such that the minority community receives the lower share of the contestable part of the public sphere under both assimilation and separation whenever* $\sigma_M > \sigma_N - \epsilon$;
(c) *the minority's share of the contestable part of the public sphere falls when it assimilates; and*
(d) *any expansion in the contestable part of the public sphere reduces the private consumption of every individual in society; furthermore, there exists* $\epsilon > 0$ *such that any such expansion reduces the minority's share of the contestable part (i.e.,* p_N*) whenever* $\sigma_M < \sigma_N + \epsilon$.

Proof See the Appendix.

Proposition 3 extends the efficiency case for assimilation (Proposition 2) to conflict over division of the public sphere. Intuitively, it focuses on the situation where ethno-religious or racial identity-based political mobilization is extensive within both communities, and the two communities are not-too-dissimilar in cohesion. In such situations, Proposition 3(a) suggests that resource loss due to conflicts may fall as a proportion of total output when N assimilates.[11] Given roughly similar community cohesion, the minority community spends a larger

[11] An off-shoot of assimilation, at least over time, may conceivably be the weakening of community cohesion within the minority, and therefore of conflicts with the majority. Kuran and Sandholm (2008) offer an evolutionary game theoretic perspective on this view. My static argument is independent of this dynamic argument. Note also that a shift to assimilation may *increase* inter-group conflicts when N is sufficiently *less cohesive*, relative to M. Then, income inequality engendered within N by assimilation may actually increase total political spending by N, expressed as a proportion of total societal income.

proportion of its income on conflict, since it suffers less from the free-rider problem due to its smaller size. Hence, since assimilation reduces the income share of the minority (Proposition 2(a)), aggregate resource loss due to conflict falls as a proportion of social output. Interestingly however, greater income of the majority community more than compensates for the free-riding effect, so that N always receives less than half the contestable part of the public sphere; assimilation reduces its share (Proposition 3(b) and (c)). If a larger part of the public sphere opens up for contestation, then all members of society increase their political spending. However, M increases its political spending proportionately *more*. Hence, such an expansion (or, equivalently, a contraction of the secular/race-blind space) leads to N faring worse in the political arena: its share of the public sphere falls (Proposition 3(d)).

Remark 1 Total resource expended on identity-related consumption conflict increases if total income accruing to *both* communities increases,[12] as must be the case under assimilation if $m > E(c_N)$ (recall Proposition 2(a and c)), but not necessarily otherwise.

Now, recalling (8) and using a log transformation of the utility function there, we have:

$$\frac{du_j^N}{dt} = \frac{1}{x_j}\frac{\partial x_j}{\partial t} + \sigma_N \ln\left(\frac{b_N}{b_M}\right) + \frac{\sigma_N t}{\left(\frac{b_N}{b_M}\right)}\frac{\partial\left(\frac{b_N}{b_M}\right)}{\partial t}. \tag{11}$$

In the light of (7) and Proposition 3 (parts (b) and (d)), (11) yields the following.

Corollary 1 *Given any triple* $< m, F^N(c), F^M(c) >$ *satisfying Assumption 1, given any* $t \in (0, 1]$, *and given any* $\sigma_N > \sigma_N^*(t)$, *there exists* $\epsilon > 0$ *such that a marginal contraction in the contestable part of the public sphere increases the welfare of every minority individual whenever* $\sigma_M \in (\sigma_N - \epsilon, \sigma_N + \epsilon)$.

Despite its aggregate conflict-reducing effects (Proposition 3), the inequality engendering effects of assimilation spill over from the space of incomes (Proposition 2) to the space of utilities.

Corollary 2 *Given any triple* $< m, F^N(c), F^M(c) >$ *satisfying Assumption 1, any* $t \in (0, 1]$, *and any* $\sigma_N > \sigma_N^*(t)$,

(a) *the welfare of every majority individual is higher under assimilation, relative to separation; and*
(b) *when* $m < \rho_N$, *the welfare of every minority individual is lower under assimilation, relative to separation.*

Proof See the Appendix.

Remark 2 Since the relative identity composition of the public sphere necessarily shifts against the minority (Proposition 3(c)), minority individuals may be worse off

[12]This follows immediately from condition (22) in the Appendix.

on assimilation even if they achieve income gains. Thus, the assumption of large identity costs relative to income gains ($m < \rho_N$) is sufficient, but not necessary, for assimilation to make *all* minority individuals worse off.

4 Discussion and Concluding Remarks

This paper has developed a parsimonious framework within which the case for assimilating minorities may be examined, and its implications for non-pecuniary social conflicts clarified. I have shown that the justification for behavioural homogenization may be deduced from (a) the productivity gains it may provide by facilitating economic interaction, and (b) the dampening effect it may have on political contestation among communities for control over the public sphere. However, these possible gains have to be balanced against the dis-equalizing consequences of integration, both within and across communities. Integration may shift the distribution of both material and symbolic goods against minorities. Second, it may expand income inequality within the minority community itself.

My analysis explains why attempts to integrate *large* minorities into majority ethno-linguistic behavioural conventions may meet with strong resistance, even if there are potential gains from such integration. To illustrate, soon after the formation of Pakistan in 1947, large-scale political conflict broke out in the eastern part of the country over attempts to make Urdu the sole official language. Consequent hardening of oppositional identities between the Urdu-speaking western part and the Bangla-speaking eastern part eventually led to civil war, genocide and the formation of Bangladesh in 1971. Similar attempts at linguistic unification sparked off decades of civil war between Sinhala-speaking Buddhists and Tamil-speaking Hindus in Sri Lanka. [13]

In explaining such conflicts, this paper can be seen as also providing functional micro-foundations for an argument for the 'rights of nationalities', with *language* as the basis for national identities. Such arguments have been used since the 19th century. They were used initially in Europe to develop a case for German and Italian unifications. Subsequently, they were deployed to justify the formation of new states from the ruins of the Czarist, Austro-Hungarian and Ottoman empires. Later, they were utilised to justify the organization of multi-ethnic, multi-religious and multi-lingual countries (such as the former Soviet Union, the former Yugoslavia, and India) along federal lines, with administrative units organized broadly on the basis of ethno-linguistic categories. The present analysis suggests that such a form of political organization, by permitting large ethno-linguistic groups to develop

[13] Montalvo and Reynal-Querol (2005, 2008) find that societies which are ethnically more *polarized*, i.e., where majority and minority communities are close in size, may be more prone to social conflicts (specifically civil wars and genocides). Easterly et al. (2006) present a similar finding in the context of mass killings. The analysis here can be seen as providing a theoretical rationalization of these empirical findings. See also footnote 17.

their own cultural-linguistic identities/conventions and organizing their economic interactions on the basis of such identities/conventions, may have served to equalize welfare both within and across communities.

However, my findings also point to a major source of potential instability in such federations. Given a history of past antagonisms, and given an overlap between linguistic and ethnic/religious/racial fault-lines, such federations need to devise methods to credibly pre-commit to keeping large segments of the public sphere, broadly interpreted, outside the scope of ethno-religious or racial identity-based political contestations. The failure to do so might generate high levels of conflict over the symbolic and normative content of the public sphere. Analogous requirements of blindness and constitutional rules apply to arbitration mechanisms for resolving conflicting demands by different regions for fiscal transfers, which are likely to become more strident and less open to compromise when inter-regional migration opportunities are restricted by ethno-linguistic fragmentation of the national labour market. Furthermore, such states run the risk of providing inefficient protection to linguistic communities too small to be viable on their own, thereby generating moral hazard problems. In India, constantly proliferating demands for the carving out of new states, especially in the north-eastern part of the country, are usually met by a combination of regional autonomy and ad-hoc fiscal transfers, which in turn incentivize new ethnicity and language based mobilizations.

These factors generate significant social losses which can potentially be reduced by integrating minorities, especially through linguistic unification. However, for such gains to actualize, assimilation costs have to be low throughout (i.e., $\overline{\rho}_N$ low). Otherwise, the segment within the minority which loses out due to assimilation (those with assimilation costs in $((1 - ny_N), \overline{\rho}_N))$ would be large, and this large segment therefore may well block attempts to assimilate it. Persistence may lead to protracted civic conflict, and the minority community may itself get split between those who wish to assimilate and those who do not.[14] One way to reduce assimilation costs might be to encourage assimilation to norms that incorporate elements from the minority culture, rather than being exclusively reflective of the majority. Gandhi wanted the national language of independent India to be Hindustani, which he conceptualized as a culturally composite language with Sanskrit as well as Arabic and Persian roots. Attempts in Europe and North America to develop a 'multicultural' syllabus in public schools, which provides positive exposure to minority expressive conventions and cultures, may also be interpreted in terms of such a project. My analysis, while sympathetic to such projects, also serves to identify their limitations. To the extent that this attempted composite diverges substantially from the majority's conventions, it imposes significant adjustment costs on the *majority*. These costs may easily exceed the gains to the minority, leading to aggregate social losses. Thus, the efficiency case for such cultural compromises is not self-evident. Nor is their political sustainability, since they are likely to generate a political

[14]Conflicts within the African-American community over 'acting White' constitute a specific example, of which Austen-Smith and Fryer (2005) provide a formalization.

backlash from majorities. Reflections of these political tensions can be perceived in conflicts over the content and organization of the public education system in Western Europe and North America over the last two decades. In India, on the other hand, the canonical formulation of the national language, Hindi, has moved increasingly closer to its Sanskrit origins and away from Arabic and Persian influences, while the converse is arguably true for the trajectory of the national language, Urdu, in Pakistan. While conscious political choices exercised through language academies certainly played a role in these developments, they are also a consequence of the cultural distance between majorities in these countries and the linguistic traditions identified with their respective minorities.

Both majorities and minorities may however have an incentive to adopt the behavioural conventions of a relatively neutral, but large, *third* community. This third community may be a supra-national entity with global presence, integration with which brings the advantage of access to a global market. Despite decolonization, languages (and cultural-behavioural conventions) of the former colonizers, especially English, French, Spanish and Portuguese, continue to be widely and officially adopted in Latin America, Africa and Asia. The results here suggest that such adoption may serve to integrate diverse and antagonistic ethno-linguistic communities within a country. Indeed, large majorities may be willing to forgo linguistic dominance over minorities *only* in favour of common assimilation to a third language that carries large benefits. Minorities may also find such assimilation more acceptable than linguistic surrender to the majority because of the cultural-historical neutrality of such a third language (which entails lower assimilation costs) and greater global scope (which increases the benefits). Thus, increased integration with global markets, including labour markets and markets for cultural production, might facilitate integration *within* individual countries, while disruption of such links might exacerbate internal group conflicts. [15] By the same token, integration with external markets may *increase* minority separatism and thereby *increase* internal conflicts when such markets deploy cultural-linguistic conventions closer to those of the minority.[16] Relatedly, internal presence of a large 'buffer' community,

[15]There is some weak cross-country evidence linking greater external openness with lower internal conflict, and it is well-known that globalization affects domestic conflict in contradictory ways through channels such as income distribution, international prices for contestable mineral resources, revenue base of the government etc. [see Barbieri and Reuveny (2005) and Magee and Massoud (2011) for recent discussions]. I thus add to this literature by highlighting an additional mechanism. Collapse of the Soviet Union and the consequent economic disruption arguably played an important role in the revival of ethno-linguistic tensions in parts of Eastern and Central Europe, as well as in many former Soviet republics. Integration into some third linguistic-cultural tradition shifts the normative issue of a just distribution of gains to a global level (see Van Parijs (2011).

[16]In recent decades, opening up of job opportunities in Saudi Arabia has led to important income gains for some sections of Indian Muslims, but has also incentivized greater adoption of Saudi Wahhabism-inspired behavioural and religious norms and expanded the influence of Arabic in expressive practices. Remittances have funded ethno-religious assertion (e.g., the building and refurbishment of denominational mosques and religious schools, lavish spending on ceremonies, withdrawal of women from the labour market, campaigns for strict observance of dress and dietary

culturally-linguistically roughly neutral, between two historically antagonistic communities, may help facilitate integration. My analysis thus highlights the possible importance of the role played by ethno-linguistic fragmentation in *reducing* civic conflict.[17]

Financial compensation may play a role in inducing minorities to integrate. Such schemes, including effective anti-discrimination or affirmative action legislation, can be modelled in my framework as an identical increase in returns from assimilation for all minority individuals (footnote 6). However, their actual working involves multiple difficulties. If transfers are made conditional on irreversible assimilation, then the majority has an incentive to renege on its commitments. This has, for example, been the case with treaties signed between the US government and various Native American nations throughout the 19th century. Second, such schemes face standard adverse selection and moral hazard problems: they may require payments to the minority in excess of its actual costs of assimilation, due to the difficulty of measuring these idiosyncratic costs (or assimilatory achievements) with any degree of confidence. As is well-known in the literatures on affirmative action programs and anti-poverty transfers, such payments may also set up perverse individual incentives that reduce minority efforts to acquire productivity enhancing skills.[18] Reinterpretation of the insights generated by these cognate literatures to the issue of optimal compensation schemes, within my framework of explicit interactions between minority linguistic integration and identity conflict, remains an open research area.

Lastly, the present paper points to the interesting possibility of non-financial compensation to minorities through expansion of the part of the public sphere closed to political competition between communities (Corollary 1). It suggests that minorities may be more open to assimilation in productivity-relevant cultural (especially linguistic) conventions when associated with measures to reduce majority ethno-religious or racial control over the symbolic and normative aspects of the public sphere. Thus, when the minority differs from the majority in both language and religion, making the majority language the sole medium of instruction

codes, etc.), and on mobilizations to organize, defend or enforce such assertion. This in turn has generated conflict and counter-mobilization. The Saudi influence is noticeable in conflicts over organized attempts to impose Wahhabism-inspired linguistic, behavioural and religious norms in Bangladesh and Pakistan as well. See, for example, Boone (2014) for a discussion in the context of Pakistan.

[17]Note that such societies are ethno-linguistically *less polarized*, and recall footnote 13. Desmet et al. (2009) and Desmet et al. (2012) develop an empirical operationalization of the idea of linguistic *distance* between communities. Consistent with my analysis, the latter contribution finds deep linguistic cleavages (which imply large assimilation costs) to be empirically better predictors of civil conflict.

[18]On affirmative action, see, for example, Holzer and Neumark (2000). Transfers conditional on assimilation efforts (e.g., participation in language classes) may also generate socially *excessive* adjustment by minorities: Bougheas et al. (2007) show how conditional anti-poverty transfers may be inefficient, yet persist indefinitely.

in schools may face less opposition if ‘bundled’ with policies to secularize an education system largely controlled by majority religious organizations. Whether such policy bundling works in general is an empirical question that deserves in-depth scrutiny.

Appendix

Proof of Proposition 1 For a majority individual, the net gain from unilaterally adopting the minority’s norms when $m_N = 0$ is at most $[n(1 - \rho_M) - m] < 0$. For a minority individual j, the net assimilation premium is:

$$P_{jN} = \left[(m + n_M)\left(1 - c_j\right) - (n - n_M)\, y_N\right]; \tag{12}$$

so that (recalling (2)), for the marginal assimilating N individual, the net gain from assimilating is:

$$Z_N(n_M) \equiv \left[(m + n_M)\left(1 - \breve{c}_N(n_M)\right) - (n - n_M)\, y_N\right]. \tag{13}$$

Complete separation (i.e., $n_M = 0$) is an equilibrium iff $Z_N(0) \leq 0$, while complete assimilation (i.e., $n_M = n$) is an equilibrium iff $Z_N(n) \geq 0$. Partial integration is an equilibrium iff, for some $n_M \in (0, n)$, $Z_N(n_M) = 0$. First notice that, since $\breve{c}_N(0) = \rho_N$ (recall (3)), (13) implies:

$$Z_N(0) = \left[m\left(1 - \rho_N\right) - n y_N\right)\right]. \tag{14}$$

Assumption 1 and (14) together imply $Z_N(0) < 0$. Hence complete separation is an equilibrium. Now recall $\breve{c}_N(n) = \overline{\rho}_N < 1$. Hence, from (13), $Z_N(n) > 0$, implying that complete assimilation by the minority community must also be an equilibrium. Noting that $Z_N(n_M)$ is continuous, it follows that the equilibrium involving complete assimilation by the minority, and that involving complete separation, are both locally stable. The proof of the claim made regarding the existence of a locally stable equilibrium involving majority acquisition of minority norms is exactly analogous. Lastly, notice that, by (13),

$$Z\prime_N(n_M) \equiv 1 + y_N - \left[\breve{c}_N(n_M) + (m + n_M)\, \breve{c}\prime_N(n_M)\right]. \tag{15}$$

From (3)–(5), $\breve{c}(n) = \overline{\rho}_N,\ \breve{c}'(n) = \frac{(\overline{\rho}_N - \rho_N)}{n\alpha_N} > 0$, and $\breve{c}''(n_M) > 0$. Then recalling $y_N \leq 1$, $Z''_N(n) < 0$ for all $n_M \in (0, n]$ and $\lim_{n_M \to 0} Z\prime_N(n_M) = (1 + y_N - \rho_N) > 0$. Recalling that $Z_N(0) < 0$, $Z_N(n) > 0$, it follows that there exists exactly one equilibrium involving partial assimilation by the minority, but this equilibrium is unstable. ∎

Proof of Proposition 2 If N universally assimilates, total output is given by:

$$Y_M = [1 - nE(c_N)]; \tag{16}$$

where $E(c_N)$ is the expected assimilation cost for an N individual. Output under separation is:

$$Y_S = \left[(1-n)^2 + n^2 y_N\right]. \tag{17}$$

M's total income under integration by N is m; whereas it is m^2 under separation. Hence, integration by the minority increases total income of the majority. Since incomes are identical within M, all its members earn more. Under assimilation by N, M's income share, using (16), is:

$$H_{MM} = \frac{m}{1-(1-m)E(c_N)}; \tag{18}$$

while under separation, using (17), it is:

$$H_{MS} = \frac{m^2}{(1-m)^2 y_N + m^2}. \tag{19}$$

Using (18) and (19), I get:

$$H_{MM} \leq H_{MS} \text{ iff } E(c_N) \leq \left(1 + y_N - \frac{y_N}{m}\right). \tag{20}$$

Now since $\rho_N < E(c_N)$, (20) implies: $H_{MM} \leq H_{MS}$ only if $\rho_N < \left(1 + y_N - \frac{y_N}{m}\right)$, which violates Assumption 1. Hence $H_{MM} > H_{MS}$. Part (a) of Proposition 2 follows.

By (16) and (17), total output is higher under assimilation iff $E(c_N) < [2m + (1-m)(1-y_N)]$. Since $E(c_N) < 1$, $m > \frac{1}{2}$, and $y_N \leq 1$, part (b) follows. Now, under assimilation, total income received by N is $n[1 - E(c_N)]$, while that under separation is $n^2 y_N$. Part (c) of Proposition 2 follows. Lastly, income received by a minority individual under assimilation is $(1 - c_{N,j})$, while that under separation is ny_N. Comparing, I get part (d) of Proposition 2. ■

Proof of Proposition 3 Let Y be the total income in society, and let $\widehat{H}_i$ be the share of political contributors in i. From (9), aggregating over the contributing population of each community, and letting $\widehat{X}^i$, $\hat{s}_i$ denote, respectively, total private consumption and population share of contributing members of i, I get the Nash equilibrium conditions: $\forall i \in \{M, N\}, \left[\widehat{X}^i = \widehat{H}_i Y - b_i = \frac{b_i \hat{s}_i}{\sigma_i t}\right]$; so that:

$$\forall i \in \{M, N\}, \left[b_i = \frac{\widehat{H}_i Y \sigma_i t}{\hat{s}_i + \sigma_i t}\right]. \tag{21}$$

Recall that individual incomes are identical within each community under separation, while they are identical within the majority community under assimilation by the minority. Hence, in these three cases, the set of contributors within a community must be the entire community. Since $\sigma_N > \sigma_N^*(t)$, the set of contributors within a community is the entire community in the remaining case as well. Thus, denoting by $H_{iS},\ \ H_{iM}$ the income share of the community i in the separated and assimilated equilibrium, respectively, from (21), the equilibrium community political contributions are:

$$b_{NS} = \frac{H_{NS} Y_S \sigma_N t}{n + \sigma_N t}, b_{MS} = \frac{H_{MS} Y_S \sigma_M t}{m + \sigma_M t}, b_{MM} = \frac{H_{MM} Y_M \sigma_M t}{m + \sigma_M t}, b_{NM} = \frac{H_{NM} Y_M \sigma_N t}{n + \sigma_N t}. \tag{22}$$

From (22), total political expenditure, expressed as a proportion of total income, is given by:

$$B_S = \left[\frac{(1 - H_{MS})}{\frac{n}{\sigma_N t} + 1} + \frac{H_{MS}}{\frac{m}{\sigma_M t} + 1} \right],\ B_M = \left[\frac{(1 - H_{MM})}{\frac{n}{\sigma_N t} + 1} + \frac{H_{MM}}{\frac{m}{\sigma_M t} + 1} \right]; \tag{23}$$

From (23), I get:

$$B_S > B_M \ \text{ iff } \ (H_{MM} - H_{MS}) \left[\frac{1}{\frac{n}{\sigma_N t} + 1} - \frac{1}{\frac{m}{\sigma_M t} + 1} \right] > 0. \tag{24}$$

Since $n < m$ and $H_{MM} > H_{MS}$ (by Proposition 2(a)), (24) implies:

$$B_S > B_M \text{ if } \sigma_N \geq \sigma_M. \tag{25}$$

Part (a) of Proposition 3 follows from (24) and (25) by continuity.

Noting (22), under separation, $\frac{b_{NS}}{b_{MS}} = \left(\frac{n y_N}{m}\right) \frac{(mn\sigma_N + n\sigma_M\sigma_N t)}{(nm\sigma_M + m\sigma_N\sigma_M t)} < 1$ if $\sigma_N \leq \sigma_M$; while, under assimilation, $\frac{b_{NM}}{b_{MM}} = \frac{(mn\sigma_N + n\sigma_M\sigma_N t)}{(nm\sigma_M + m\sigma_N\sigma_M t)} < 1$ if assimilation costs are 0. Hence, N's political expenditure must be lower than M's in either case when $\sigma_N \leq \sigma_M$. Part (b) of Proposition 3 follows by continuity.

From (22),

$$\frac{b_{NS}}{b_{MS}} = \left(\frac{H_{NS}}{H_{MS}} \right) \frac{\sigma_N (m + \sigma_M t)}{\sigma_M (n + \sigma_N t)}, \frac{b_{NM}}{b_{MM}} = \left(\frac{H_{NM}}{H_{MM}} \right) \frac{\sigma_N (m + \sigma_M t)}{\sigma_M (n + \sigma_N t)} \tag{26}$$

Since M's income share is higher under assimilation (Proposition 2(a)), (26) implies part (c).

From (22), any increase in the contestable part of the public sphere (i.e., in t) reduces the private consumption of all members of society. Using (26), I also have:

$$\frac{d\left(\frac{b_N}{b_M}\right)}{dt} = \left(\frac{H_N}{H_M}\right)\left[\frac{\sigma_N {\sigma_M}^2 (n+\sigma_n t) - \sigma_M {\sigma_N}^2 (m+\sigma_m t)}{{\sigma_M}^2 (n+\sigma_n t)^2}\right]$$
$$= \sigma_N \sigma_M \left(\frac{H_N}{H_M}\right)\left[\frac{\sigma_M n - \sigma_N m}{{\sigma_M}^2 (n+\sigma_n t)^2}\right]. \tag{27}$$

From (27), $\frac{d\left(\frac{b_N}{b_M}\right)}{dt} < 0$ if $\sigma_N \geq \sigma_M$. Part (d) follows from (27) by continuity. ■

Proof of Corollary 2 By Proposition 2(a), assimilation increases the income of every M individual. Then, by (9), private consumption of every M individual must rise. By Proposition 3(c), M's share of the contestable part of the public sphere rises when N assimilates. Part (a) of Corollary 2 follows from (8). Now, when $m < \rho_N$, assimilation lowers income of every N individual (Proposition 2(d)). Suppose private consumption increases for some N individual. Then, by (9), total minority political expenditure must rise, which in turn implies that private consumption must rise for *every* N individual, so that total N income must rise: a contradiction. Hence, private consumption must fall for every N individual. Part (b) of Corollary 2 follows from (8) and Proposition 3(c). ■

References

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics, 94*(4), 749–775.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics, 115*(3), 715–53.

Austen-Smith, D., & Fryer, R. G., Jr. (2005). An economic analysis of 'acting White'". *Quarterly Journal of Economics, 120*(2), 551–583.

Bakshi, D., & Dasgupta, I. (2017, forthcoming). A model of dynamic conflict in ethnocracies. *Defence and Peace Economics*. doi: 10.1080/10242694.2015.1092204

Barbieri, K., & Reuveny, R. (2005). Economic globalization and civil war. *Journal of Politics, 67*(4), 1228–1247.

Bergstrom, T. C., Blume, L., & Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics, 29*, 25–49.

Bhattacharya, S., Deb, J., & Kundu, T. (2015). Mobility and conflict. *American Economic Journal: Microeconomics, 7*(1), 281–319.

Boone, J. (2014, January 15). The saints go marching out as the face of Islam hardens in Pakistan. *The Guardian*, *UK*. www.theguardian.com

Bougheas, S., Dasgupta, I., & Morrissey, O. (2007). Tough love or unconditional charity? *Oxford Economic Papers, 59*(4), 561–582.

Bowles, S., Sethi, R., & Loury, G. (2014). Group inequality. *Journal of the European Economic Association, 12*(1), 129–152.

Cornes, R. C., & Schweinberger, A. G. (1996). Free riding and the inefficiency of the private production of pure public goods. *Canadian Journal of Economics, 29*(1), 70–91.
Dasgupta, I. (2009). 'Living' wage, class conflict and ethnic strife. *Journal of Economic Behavior and Organization, 72*(2), 750–765.
Dasgupta, I., & Kanbur, R. (2005a). Community and anti-poverty targeting. *Journal of Economic Inequality, 3*(3), 281–302.
Dasgupta, I., & Kanbur, R. (2005b). Bridging communal divides: Separation, patronage, integration. In C. Barrett (Ed.), *The social economics of poverty: On identities, groups, communities and networks* (pp. 146–170). London: Routledge.
Dasgupta, I., & Kanbur, R. (2007). Community and class antagonism. *Journal of Public Economics, 91*(9), 1816–1842.
Dasgupta, I., & Kanbur, R. (2011). Does philanthropy reduce inequality? *Journal of Economic Inequality, 9*(1), 1–21.
Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2012). The political economy of linguistic cleavages. *Journal of Development Economics, 97*, 322–338.
Desmet, K., Ortuño-Ortín, I., & Weber, S. (2009). Linguistic diversity and redistribution. *Journal of the European Economic Association, 7*(6), 1291–1318.
Easterly, W., Gatti, R., & Kurlat, S. (2006). Development, democracy and mass killings. *Journal of Economic Growth, 11*(2), 129–56.
Esteban, J., & Ray, D. (2008). On the salience of ethnic conflict. *American Economic Review, 98*, 2185–2202.
Esteban, J., & Ray, D. (2011). A model of ethnic conflict. *Journal of the European Economic Association, 9*(3), 496–521.
Fredrickson, G. M. (1999). Models of American ethnic relations: A historical perspective". In D. A. Prentice & D. T. Miller (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp. 23–34). New York: Russell Sage.
Hobsbawm, E. J. (1992). *Nations and nationalism since 1780* (2nd ed.). New York: Cambridge University Press.
Hobsbawm, E. J., & Ranger, T. (Eds.). (1983). *The invention of tradition*. New York: Cambridge University Press.
Holzer, H., & Neumark, D. (2000). Assessing affirmative action. *Journal of Economic Literature, 38*(3), 483–568.
Kónya, I. (2005). Minorities and majorities: A dynamic model of assimilation. *Canadian Journal of Economics, 38*(4), 1431–1452.
Kuran, T., & Sandholm, W. (2008). Cultural integration and its discontents. *Review of Economic Studies, 75*, 201–228.
Lazear, E. (1999). Culture and language. *Journal of Political Economy, 107*(6), S95–S126.
Lewis, D. (1969). *Convention: A philosophical study*. Cambridge: Cambridge University Press.
Li, Q. (2013). Language and urban labor market segmentation: Theory and evidence. *Journal of Urban Economics, 74*, 27–46.
Magee, C., & Massoud, T. (2011). Openness and internal conflict. *Journal of Peace Research, 48*(1), 59–72.
Montalvo, J., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict and civil wars. *American Economic Review, 95*(3), 796–816.
Montalvo, J., & Reynal-Querol, M. (2008). Discrete polarization with an application to the determinants of genocides. *Economic Journal, 118*(533), 1835–65.
Ortega, J., & Tangerås, T. (2008). Unilingual versus bilingual education: A political economy analysis. *Journal of the European Economic Association, 6*(5), 1078–1108.
Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
Van Parijs, P. (2011). *Linguistic justice for Europe and the world*. New York: Oxford University Press.

International Carbon Trade and National Taxes: Distributional Impacts of Double Regulation

Thomas Eichner and Rüdiger Pethig

1 Introduction

We envisage an international agreement on mitigating climate change like the Kyoto Protocol or a possible Post-Kyoto agreement, in which a group of countries commits to reduce carbon emissions relative to their baseline emissions. Each country in that group is assumed to observe an emissions limit, called national emissions cap. Varying the distribution of national emissions caps under the constraint of keeping constant the total emissions limit for the entire group is known to have consequences for the countries' welfare. Since the national emissions cap is a valuable asset for the individual country, the larger is that asset the better off the country tends to be at the expense of the other countries. To put it differently, each country's share in the burden of implementing the group emissions cap tends to be the smaller the larger is its national emissions cap. Although that observation is well understood, in general, the precise impact of any given distribution of caps on the distribution of burdens depends on the institutional design of emissions control. In particular, the distribution of burdens depends on whether or not:

(a) The group of countries operates a joint emissions trading scheme (ETS);
(b) The ETS covers only part of each country's economy (which we will call ETS sector);

T. Eichner (✉)
Department of Economics, University of Hagen, Universitätsstr. 41, 58097 Hagen, Germany
e-mail: thomas.eichner@fernuni-hagen.de

R. Pethig
Department of Economics, University of Siegen, Unteres Schloss 3, 57072 Siegen, Germany
e-mail: pethig@vwl.wiwi.uni-siegen.de

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_13

(c) The individual country regulates carbon emissions not only in its non-ETS sector, which is a necessary condition for cost-effective regulation, but also in its ETS sector, which constitutes potentially harmful *double regulation*.

In an analytical framework with the features (a)–(c), the present paper aims to analyze the impact of double regulation via ETS and overlapping national emissions taxes on the distribution of national welfares. That analysis is empirically relevant, because the arrangements (a)–(c) of emissions control characterize the European Union's current approach to fulfilling its Kyoto emissions reduction obligations. In fact, the EU has complemented its member states' national emissions control by a joint ETS in 2005 (EU 2003) which covers only part of each member state's economy. In their non-ETS sectors, national governments are responsible for curbing emissions by means of domestic policies. There also (pre-)exist various national regulatory policies, notably energy taxes, in the ETS sector overlapping with the ETS (Johnstone 2003; Sorell and Sijm 2003; IEA 2011; Lehmann 2012). The focus of the paper on the distributional impact of double regulation may become even more relevant, if in the near future—which is conceivable, though not yet clear at present—some Post-Kyoto agreement should enter into force which extends the EU-type approach to carbon reduction beyond the EU to some (or even many) non-EU countries.

For any given distribution of national emissions caps, emissions taxes in the countries' ETS sectors impact on national welfares in that they affect the equilibrium permit price and thus each country's export or import of permits. It is therefore the overlapping tax as well as the distribution of national emissions caps that determine the distributional incidence of mixed policies. As an implication, the calculation of national net burdens of mixed policies is not correct unless the interaction of both instruments is accounted for. To address this issue we will consider variations in double regulation and investigate their *net* effect on national welfares resulting from an integrated account of the *partial* welfare effects of both instruments that may point either in the same or in opposite directions. Policies using a tax without ETS or an ETS without a tax are included as limiting cases as illustrated in Table 1.

To keep the focus on distribution as clear as possible, we assume constant throughout the paper the emissions limit for the entire group of countries. First, we analyze cost-ineffective policies, i.e. double regulation with overlapping taxes whose rates differ across countries. Proposition 1 shows that for every cost-ineffective mixed policy there exists a (cost-effective) ETS-only policy providing all countries with a level of welfare higher than that in the mixed policy by some

Table 1 Emissions control in a two-sector economy

	Emissions control in the ETS sector via		
	ETS	ETS and sectoral tax	Sectoral tax
Emissions control in the non-ETS sector via sectoral tax	1	2	3

uniform percentage rate. In the rest of the paper we restrict attention to cost-effective policies. For the study of those policies we introduce the simplifying assumptions[1,2]

- that in their non-ETS sectors national governments effectively control emissions through a domestic sectoral emissions tax;
- that the rate of the emissions tax overlapping with the ETS sector is uniform across countries (which renders the double regulation cost effective)[3];
- and that all countries choose their permit cap efficiently, as to equalize marginal abatement cost across the ETS sector and the non-ETS sector.

Having characterized the associated class of cost-effective hybrid policies we first look at the limiting case of a tax-only policy (box 3 in Table 1) as a useful benchmark and show that the associated equilibrium allocation is unaffected by the introduction of an ETS and its partition into national caps (Proposition 2). Next we characterize the *distributional* consequences of mixed policies (box 2 in Table 1) by showing how a country's welfare varies in response either to changes in the emissions tax rate in the ETS sectors or to changes in its national emissions cap (with compensating changes in the caps of all other countries). Due to the interdependence of markets, the distributional effects of policy changes turn out to be non-monotone, in general, and hence not easy to characterize (Proposition 3). Next, we establish an equivalence result (Proposition 4) stating that for every mixed policy (box 2 in Table 1) there exists an ETS-only policy (box 1 in Table 1) which provides all countries with the same level of welfare and the same allocation as the mixed policy. It is also possible to specify how the national caps in the equivalent ETS-only policy deviate from the caps assigned to the countries in the actually prevailing mixed policy (Proposition 5).

Making use of our analytical findings we finally propose two measures for the distributional incidence of emissions control. The first measure is non-monetary taking advantage of both the equivalence result and the benchmark property of the tax-only policy. This measure allows to identify winners and losers of mixed policies relative to the tax-only policy. In the spirit of the welfare measure of equivalent variation the second measure consists of a monetary transfer a country needs to pay or receive in order to be indifferent between some given mixed policy and the tax-only benchmark policy.

Most of the literature on hybrid carbon emissions control deals with allocative distortions of existing policies and/or with issues of policy design for allocative

[1] Here we follow Eichner and Pethig (2009) who established conditions under which the policy mix is cost effective for the group of countries. They show, in particular, that the emissions tax can be fixed at different levels without compromising cost effectiveness if the overlapping tax is uniform across countries (and if some other qualifications are met). For more details see also Sect. 2.

[2] The relation between cost effectiveness and Pareto efficiency has been clarified by Chichilnisky and Heal (1994) and Shiell (2003).

[3] Nordhaus (2006) brought forward arguments in favor of an internationally harmonized emissions tax.

efficiency, e.g. Bovenberg and de Mooij (1994), Babiker et al. (2003), Bento and Jacobsen (2007), Böhringer et al. (2008), Mandell (2008) and Eichner and Pethig (2009). Only a few studies address the international distribution of national welfares and burdens. The issue of *equitable* burden sharing has been studied e.g. by Phylipsen et al. (1998) and Marklund and Samakovlis (2007). Caplan et al. (2003) study an international ETS with redistributive resource transfers. To be more specific, countries choose emissions caps in a Nash fashion and a Global Environmental Facility operates a transfer scheme. They show that the combination of emissions trading and transfer mechanism yields efficiency. Yet our focus is not on efficiency, equity or fairness but rather on the *positive* analysis of the distributional impact of mixed policies. There is an applied literature of numerical analysis in large-scale CGE models in which some distributional issues are investigated although not in a systematic *analytical* way. For example, Böhringer et al. (2008) consider a group of countries operating an ETS and they calculate how burdens change when an individual country successively raises the rate of the emissions tax in its ETS sector. Peterson and Klepper (2007) compare a harmonized international carbon tax to an ETS with different allocation rules for the emissions caps without considering the issue of overlapping instruments. Hence the distributional incidence of mixed policies appears to be under-researched.

The paper is organized as follows: Sect. 2 sets up the model. Section 3 characterizes cost-effective equilibria and specifies the distributional impacts of cost-*in*effective policies. Section 4 establishes the tax-only policy as a benchmark and analyzes the welfare effects of changes in policy parameters. Section 5 points out that every mixed cost-effective policy can be transformed into a cost-effective ETS-only policy without changing the associated welfare distribution. Section 6 suggests two measures of the distributional impact of mixed policies and Sect. 7 concludes.

2 The Model

We consider an economy of n countries that are open to the rest of the world and operate a joint ETS. Each country's economy consists of two sectors: One sector is covered by the ETS, called the ETS sector, and we refer to the rest of the economy as the non-ETS sector. The non-ETS sector of country $i = 1, \ldots, n$ uses the fossil fuel input e_{xi} to produce the output $x_{si} = X^i(e_{xi})$. Likewise, the ETS sector uses the fossil fuel input e_{yi} to produce the output $y_{si} = Y^i(e_{yi})$.[4] All fossil fuel is assumed to be imported from the world market at the fixed price p_e. The energy costs of the firms in country i's ETS sector are $(p_e + t_{yi})e_{yi}$, if country i levies an energy tax at rate t_{yi}. We consider that tax as a tax on carbon emissions because the release of CO_2 is approximately proportional to the amount of fossil fuel burned. The imports of

[4] The index s stands for the supply of good X and Y, respectively.

fossil fuel are (mainly) paid for by exporting good Y at the world market price p_y.[5] Good X^i is traded on a domestic market at price p_{xi} and the corresponding market clearance condition is

$$x_{si} = x_i \qquad \text{for } i = 1, \ldots, n, \tag{1}$$

where x_i is the domestic demand for good X^i. Given the overall emissions cap $\bar{c}$ for the group of countries and some partition $(c_1, \ldots, c_n)$ of $\bar{c}$ (as outlined in the Introduction) the government of each country i chooses the permit budget $c_{yi} \in [0, c_i]$. It issues and hands over to its ETS sector for free[6] the amount c_{yi} of emissions permits which can then be traded at price π_e among all firms in the ETS sectors of all countries. The condition for equilibrium on that permit market is

$$\sum_j c_{yj} = \sum_j e_{yj}. \tag{2}$$

Each country also levies an emissions tax in its non-ETS sector whose rate t_{xi} is assumed to be chosen as to satisfy

$$c_i - c_{yi} = e_{xi} \qquad \text{for } i = 1, \ldots, n. \tag{3}$$

Summing (3) over i and invoking (2) shows immediately that the overall emissions cap $\bar{c}$ is met: $\sum_j c_j = \sum_j (e_{xj} + e_{yj}) = \bar{c}$.

The representative consumer of country i derives utility $U^i(x_i, y_i)$ from consuming the amounts x_i and y_i of the goods X and Y, respectively. Her income is $z_i := g_{xi} + g_{yi} + t_{xi}e_{xi} + t_{yi}e_{yi}$. That income consists of transferred profits $g_{xi} := p_{xi}x_{si} - (p_e + t_{xi})e_{xi}$ and $g_{yi} := p_y y_{si} - \pi_e(e_{yi} - c_{yi}) - (p_e + t_{yi})e_{yi}$ and of recycled tax revenues $t_{xi}e_{xi}$ and $t_{yi}e_{yi}$. The consumer spends her income on the goods X and Y and hence observes the budget equation

$$z_i = p_{xi}x_i + p_y y_i. \tag{4}$$

The definitions of z_i, g_{xi} and g_{yi} combined with (1) and (4) yield country i's trade balance

$$p_y(y_{si} - y_i) + \pi_e(c_{yi} - e_{yi}) - p_e(e_{xi} + e_{yi}) = 0. \tag{5}$$

[5]Part of the import bill may also be paid for by revenues from exporting permits. However, if permits are imported, the import of both fossil fuel and permits needs to be paid for by revenues from exports of good Y. See the trade balance equation (5) below.

[6]At the high level of abstraction in the present analysis, free allocation and auctioning of emissions permits are equivalent allocation procedures. For an analysis where the allocation rule matters see Rosendahl (2008).

In the n-country economy described above a policy consists of a choice of instruments $\mathbf{t}_x := (t_{x1}, \ldots, t_{xn}) \in \mathbb{R}^n_+, \mathbf{t}_y := (t_{y1}, \ldots, t_{yn}) \in \mathbb{R}^n_+, \mathbf{c} := (c_1, \ldots, c_n) \in C := \left\{\mathbf{c} \in \mathbb{R}^n_+ \;\middle|\; \sum_j c_j = \bar{c}\right\}$ and $\mathbf{c}_y(\mathbf{c}) := \left[c_{y1}(\mathbf{c}), \ldots, c_{yn}(\mathbf{c})\right] \in C_y(\mathbf{c}) := [0, c_1] \times [0, c_2] \times \ldots \times [0, c_n]$. With $\mathbf{c} \in C$ and $\mathbf{c}_y(\mathbf{c}) \in C_y(\mathbf{c})$ being fixed, the emissions ceiling (3) for country i's non-ETS sector is also determined. As noted above, the sectoral emissions cap $c_i - c_{yi}$ is implemented through an appropriate choice of t_{xi}. Hence if we take $\mathbf{c}, \mathbf{c}_y(\mathbf{c})$ and $\mathbf{t}_y$ as policy decision variables, as we will do, the tax rates $\mathbf{t}_x$ are endogenous variables rather than independent policy parameters.

Having introduced the necessary notation and described the structure of the model we now define the competitive equilibrium as follows:
Let the world market prices p_e *and* p_y*, and some policy* $[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]$ *be given. The prices*[7] $\pi_e, \mathbf{p}_x$*, the tax rates* $\mathbf{t}_x$*, and the allocation* $(\mathbf{x}_s, \mathbf{e}_x, \mathbf{y}_s, \mathbf{e}_y, \mathbf{x}, \mathbf{y})$ *constitute a competitive equilibrium of the n-country economy, if (1), (2) and (3) hold and if for* $i = 1, \ldots, n$[8]:

- (x_{si}, e_{xi}) *satisfies* $e_{xi} = \text{argmax}\left[p_{xi}X^i(\tilde{e}_{xi}) - (p_e + t_{xi})\tilde{e}_{xi}\right]$ *and* $x_{si} = X^i(e_{xi})$,
- (y_{si}, e_{yi}) *satisfies* $e_{yi} = \text{argmax}\left[p_y Y^i(\tilde{e}_{yi}) - \pi_e(\tilde{e}_{yi} - c_{yi}) - (p_e + t_{yi})\tilde{e}_{yi}\right]$ *and* $y_{si} = Y^i(e_{yi})$,
- (x_i, y_i) *satisfies* $(x_i, y_i) = \text{argmax}\left[U^i(\tilde{x}_i, \tilde{y}_i)\ s.t.\ (4)\right]$.

Assuming that production functions are concave and utility functions are quasi-concave, it can be shown that for given p_e, p_y and $[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]$ (with appropriate upper bounds on the tax rates $\mathbf{t}_y$) a competitive equilibrium exists and is unique. However, such equilibria are not cost effective, in general, and the equilibrium distribution of welfares will crucially depend on the policy $[\mathbf{c}, \mathbf{c}_y, (\mathbf{c}), \mathbf{t}_y]$ chosen. To obtain a clear focus on distribution, we disentangle distribution from allocative inefficiency in the next section.

3 Distributional Impacts of Cost-Ineffective Policies

Eichner and Pethig (2009) show that cost effectiveness is attained if and only if

$$t_{xi} = t_x \quad \text{and} \quad t_{yi} = t_y \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad t_x = \pi_e + t_y. \tag{6}$$

According to (6), cost effectiveness requires marginal abatement costs, and hence producer prices of emissions, to be the same across sectors and countries. For every $\mathbf{c} \in C$ there is one and only one vector of national permit caps, denoted

[7]Throughout the paper bold letters denote row vectors, e.g. $\mathbf{p}_x := (p_{x1}, \ldots, p_{xn})$ or $\mathbf{y}_s := (y_{s1}, \ldots, y_{sn})$.

[8]The variables in the next three lines that are marked by a wiggle (like "$\tilde{e}_{xi}$") are meant to vary over $\mathbb{R}_+$. In contrast, when there is no wiggle we deal with an equilibrium value of the respective variable.

$\mathbf{c}_y(\mathbf{c}) = \mathbf{c}_y^*(\mathbf{c})$, that secures the equalization of marginal abatement costs. Hence the permit caps are no independent policy parameters anymore when policies are required to be cost effective. Unfortunately, the policy currently applied in the EU is cost *in*effective for various reasons, and it is unlikely that cost effectiveness will be attained in the event of more countries joining an EU-type policy. Two major reasons for inefficiency are inefficiently fixed permit caps, $\mathbf{c}_y(\mathbf{c}) \neq \mathbf{c}_y^*(\mathbf{c})$, and tax rates $\mathbf{t}_y = (t_{y1}, \ldots, t_{yn})$ that differ across countries.[9] These constellations of policy parameters generate not only excess costs but also distributional effects. Since our focus in the present paper is on distribution, we now aim to decompose the total effect of cost-ineffective policies into an efficiency effect and a distributional effect which will then enable us to separate the distributional effects from the overall impacts of mixed policies.

To carry out that decomposition it is convenient to introduce the following notation. We denote by D^n the set of cost-*in*effective policies, by D^e the set of those cost-effective ETS-only policies ($\mathbf{t}_y = 0$) and write $u_i[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]$ for the welfare of country i in the equilibrium associated with policy $[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]$. We restrict attention to policies $[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]$, $\mathbf{c} \in C$, for which an equilibrium exists.

Proposition 1 *There exist functions* $Z : D^n \longrightarrow D^e$ *and* $B : D^n \longrightarrow]1, \infty[$ *such that*

$$u_i\left[Z[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]\right] = B\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \cdot u_i\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \quad \text{for } i = 1, \ldots, n. \tag{7}$$

Proof For $\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \in D^n$ define

$$\rho_i\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] := \frac{u_i\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]}{u_1\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]} \quad \text{for } i = 1, \ldots, n$$

and define the policy $\left[\hat{\mathbf{c}}, \mathbf{c}_y^*(\hat{\mathbf{c}}), \mathbf{0}\right] \in D^e$ by the equations

$$u_i\left[\hat{\mathbf{c}}, \mathbf{c}_y^*(\hat{\mathbf{c}}), \mathbf{0}\right] = \rho_i\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \cdot u_1\left[\hat{\mathbf{c}}, \mathbf{c}_y^*(\hat{\mathbf{c}}), \mathbf{0}\right] \quad \text{for } i = 1, \ldots, n. \tag{8}$$

A distribution of national caps $\hat{\mathbf{c}} \in C$ satisfying (8) clearly exists and is located on the welfare possibility frontier generated by the set of welfare distributions $\mathbf{u}\left[\mathbf{c}, \mathbf{c}_y^*(\mathbf{c}), \mathbf{0}\right]$ with $\left[\mathbf{c}, \mathbf{c}_y^*(\mathbf{c}), \mathbf{0}\right] \in D^e$. Hence (8) defines the mapping $Z : D^n \longrightarrow D^e$ such that $\left(\hat{\mathbf{c}}, \mathbf{c}_y^*, \mathbf{0}\right) = Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]$. From the definition of $\rho_i\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]$ above

[9] Since the national tax rates $\mathbf{t}_y$ are not uniform in the EU, it is not second best, in general, to choose the allocation of national permit caps $\mathbf{c}_y(\mathbf{c})$, such that marginal abatement costs are the same across sectors and countries, as is optimal in case of cost-effective policies (Eichner and Pethig 2010). It is unlikely that the permit caps $\mathbf{c}_y(\mathbf{c})$ laid down in the national allocation plans of all member states are the second-best permit caps because there are no indications that the governments of the EU member states have appropriately accounted for the preexisting tax rates $\mathbf{t}_y$ in calculating those caps.

combined with (8) and $\left(\hat{\mathbf{c}}, \mathbf{c}_y^*, \mathbf{0}\right) = Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]$ follows

$$\frac{u_i\left[Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]\right]}{u_i\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]} = \frac{u_j\left[Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]\right]}{u_j\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]} =: B\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \quad \text{for } i, = 1, \ldots, n.$$

It remains to show that the ratio of utilities that we have defined above as $B\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]$ is greater than one. Since the equilibrium associated to $[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y]$ is not cost effective (because the tax rates t_{y_i} are presupposed to differ across countries) the welfare distribution of that equilibrium is clearly located below the welfare distribution frontier for the policies $\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \in D^e$. As a consequence, $B\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] > 1$. ■

The message of Proposition 1 is straightforward. For each cost-*in*effective policy $\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \in D^n$ there is a cost-effective policy without overlapping taxes, $Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] \in D^e$, such that the equilibrium welfare distribution of the former is linked to the latter via the equation

$$\mathbf{u}\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right] = \frac{\mathbf{u}\left[Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]\right]}{B\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]}.$$

In that way we are able to specify the distributional impact of the cost-*in*effective policy $Z\left[\mathbf{c}, \mathbf{c}_y(\mathbf{c}), \mathbf{t}_y\right]$. In other words, Proposition 1 allows us to restrict our attention to comparisons of cost-effective policies in the remainder of the paper.

4 Distributional Impacts of Variations in Cost-Effective Carbon Control

Recall our observation at the beginning of the previous section that if policies are cost effective, permit caps are no independent policy parameters anymore. It suffices, therefore, to describe cost-effective policies simply by $(\mathbf{c}, t_y)$, where $\mathbf{c} \in C$ and $t_y \in \mathbb{R}_+$. For convenience of notation, we will write $\pi_e(\mathbf{c}, t_y)$, $p_{xi}(\mathbf{c}, t_y)$, $x_i(\mathbf{c}, t_y)$ etc. when referring to the values of variables belonging to the cost-effective competitive equilibrium under policy $(\mathbf{c}, t_y)$. Moreover, we will denote the entire equilibrium as $E(\mathbf{c}, t_y) := \left[P(\mathbf{c}, t_y), A(\mathbf{c}, t_y)\right]$, where $P(\mathbf{c}, t_y) := [\pi_e(\mathbf{c}, t_y), \mathbf{p}_x(\mathbf{c}, t_y)]$ are the equilibrium prices and where $A(\mathbf{c}, t_y) := \left[\mathbf{x}_s(\mathbf{c}, t_y), \mathbf{e}_x(\mathbf{c}, t_y), \mathbf{x}(\mathbf{c}, t_y), \mathbf{y}_s(\mathbf{c}, t_y), \mathbf{e}_y(\mathbf{c}, t_y), \mathbf{y}(\mathbf{c}, t_y)\right]$ is the equilibrium allocation for the group of countries. If we would allow for sign-unconstrained permit prices π_e, an equilibrium would exist for all policies $(\mathbf{c}, t_y)$, $\mathbf{c} \in C$, $t_y \in \mathbb{R}$. However, we rule out $\pi_e < 0$ and therefore restrict tax rates to the set $T(\mathbf{c}) := \left\{t_y \geq 0 \mid \pi_e(\mathbf{c}, t_y) \geq 0\right\}$.

4.1 Some Basic Relations Between ETS-Only Policies, Tax-Only Policies and Mixed Policies

Before we turn to analyzing the impact of policy variations, it is useful to reveal some basic relations between the classes of (cost-effective) policies referred to in the boxes 1, 2 and 3 of Table 1. Suppose first, there is no ETS, but sectoral taxes are levied, instead, that are uniform across sectors and countries ($t_{yi} = t_{xi} = t_y$ for all i). Denote such tax-only policies as policies $(0, t_y)$. In order to secure the group emissions cap $\bar{c}$, a particular tax rate $\bar{t}_y$ needs to be chosen, defined by $\sum_j \left[e_{xj}(0, \bar{t}_y) + e_{yj}(0, \bar{t}_y)\right] = \sum_j c_j$. The policy $(0, \bar{t}_y)$ is related to policies in the boxes 1 and 2 of Table 1 as follows:

Proposition 2

(i) $A(\mathbf{c}, 0) = A(0, \bar{t}_y)$, *if and only if* $\mathbf{c} = \hat{\mathbf{c}} := \mathbf{e}_x(0, \bar{t}_y) + \mathbf{e}_y(0, \bar{t}_y)$.

(ii) $\mathrm{Max}_{t_y} \left\{t_y \geq 0 \;\middle|\; \pi_e(\mathbf{c}, t_y) \geq 0\right\} = \bar{t}_y$ *for all* $\mathbf{c} \in C$, *if the income elasticity of both consumer goods is positive.*

(iii) $A(\mathbf{c}, \bar{t}_y) = A(\mathbf{c}', \bar{t}_y)$ *for all* $\mathbf{c}, \mathbf{c}' \in C$.

Proof

Ad (i). Sufficiency. Consider the policy $(\mathbf{c}, 0)$, with $\mathbf{c} = \hat{\mathbf{c}}$. In order to construct the equilibrium $E(\hat{\mathbf{c}}, 0)$, suppose tentatively that $\pi_e(\hat{\mathbf{c}}, 0) = \bar{t}_y$ and $p_{xi}(\hat{\mathbf{c}}, 0) = p_{xi}(0, \bar{t}_y)$, all i. Given these price assignments, the production allocations are the same under both policies and so are the consumer's optimal consumption plans because the consumer prices are unchanged. Hence $A(\hat{\mathbf{c}}, 0) = A(0, \bar{t}_y)$.

Necessity. Suppose that $A(\hat{\mathbf{c}}, 0) = A(0, \bar{t}_y)$. Necessary conditions for identical production allocations are $\pi_e(\hat{\mathbf{c}}, 0) = \bar{t}_y$ and $p_{xi}(\hat{\mathbf{c}}, 0) = p_{xi}(0, \bar{t}_y)$. In case of policy $(\hat{\mathbf{c}}, 0)$ the consumer's income is $z_i(\hat{\mathbf{c}}, 0) = g_{xi}(\hat{\mathbf{c}}, 0) + g_{yi}(\hat{\mathbf{c}}, 0) + t_{xi}(\hat{\mathbf{c}}, 0) + e_{xi}(\hat{\mathbf{c}}, 0)$ or[10]

$$z_i(\hat{\mathbf{c}}, 0) = p_{xi}(\hat{\mathbf{c}}, 0) x_{si}(\hat{\mathbf{c}}, 0) + y_{si}(\hat{\mathbf{c}}, 0) - \left[e_{xi}(\hat{\mathbf{c}}, 0) + e_{yi}(\hat{\mathbf{c}}, 0)\right] p_e - \left[e_{yi}(\hat{\mathbf{c}}, 0) - \hat{c}_{yi}\right] \pi_e(\hat{\mathbf{c}}, 0). \tag{9}$$

In case of policy $(0, \bar{t}_y)$ the consumer's income is $z_i(0, \bar{t}_y) = g_{xi}(0, \bar{t}_y) + g_{yi}(0, \bar{t}_y) + \bar{t}_y \left[e_{xi}(0, \bar{t}_y) + e_{yi}(0, \bar{t}_y)\right]$ or

$$z_i(0, \bar{t}_y) = p_{xi}(0, \bar{t}_y) x_{si}(0, \bar{t}_y) + y_{si}(0, \bar{t}_y) - \left[e_{xi}(0, \bar{t}_y) + e_{yi}(0, \bar{t}_y)\right] p_e. \tag{10}$$

[10] Good Y is chosen as numeraire.

Since the equilibrium values of p_{xi}, x_{si}, y_{si}, e_{xi} and e_{yi} are the same under both policies we infer from (9) and (10)

$$z_i(0, \bar{t}_y) - z_i(\hat{\mathbf{c}}, 0) = \left[e_{yi}(0, \bar{t}_y) - \hat{c}_{yi}\right] \bar{t}_y. \tag{11}$$

Invoke $\hat{c}_{yi} = \hat{c}_i - e_{xi}(\hat{\mathbf{c}}, 0)$ from (3) to turn (11) into

$$z_i(0, \bar{t}_y) - z_i(\hat{\mathbf{c}}, 0) = \left[e_{xi}(0, \bar{t}_y) + e_{yi}(0, \bar{t}_y) - \hat{c}_i\right] \bar{t}_y, \tag{12}$$

which is presupposed to be non-zero for some countries. For these countries the consumption $\left[x(0, \bar{t}_y), y(0, \bar{t}_y)\right]$ is not optimal under policy $(\hat{\mathbf{c}}, 0)$. Hence $A(\hat{\mathbf{c}}, 0) \neq A(0, \bar{t}_y)$.

Ad (ii). Suppose not. Then there is $E(\mathbf{c}, \tilde{t}_y)$, $\tilde{t}_y \neq \bar{t}_y$ and $\pi_e(\mathbf{c}, \tilde{t}_y) = 0$. Consider first $\tilde{t}_y > \bar{t}_y$. Since $\Delta t_y > 0$ leads to $\Delta y_{si} < 0$ and $\Delta e_{yi} < 0$ for all i, permit market equilibrium requires $\Delta e_{xj} > 0$ for some $j = 1, \ldots, n$, which in turn implies $\Delta p_{xj} > 0$ and $\Delta x_{sj} = \Delta x_j > 0$.

(a) Suppose first that $\Delta e_{xi} + \Delta e_{yi} = 0$ for all i. In that case $\Delta y_j = \Delta y_{sj} - \left(\Delta e_{xj} + \Delta e_{yj}\right) p_e < 0$ because $\Delta e_{xj} + \Delta e_{yj} = 0$ and $\Delta y_{sj} < 0$. (Recall also that $\pi_e(\mathbf{c}, \bar{t}_y) = \pi_e(\mathbf{c}, \tilde{t}_y) = \Delta \pi_e = 0$.) In sum, we have $\Delta y_j < 0$, $\Delta x_j > 0$, $\Delta p_{xj} > 0$ (and $\Delta p_y \equiv 0$). Higher consumption of good X whose price has increased and lower consumption of good Y with unchanged price cannot be optimal since the income elasticity of both goods is positive by assumption. Hence $\Delta e_{xi} + \Delta e_{yi} = 0$ for all i is not compatible with $E(\mathbf{c}, \tilde{t}_y)$.

(b) Suppose next that $\Delta e_{xi} + \Delta e_{yi} \neq 0$ for some i. In that case there must be at least one country j for which $\Delta e_{xj} + \Delta e_{yj} > 0$ holds and hence $\Delta y_j < 0$. The argument of the previous paragraph carries over.

In an analogous way one can show that the equilibrium under the policy $(\mathbf{c}, \tilde{t}_y)$ with $\tilde{t}_y < \bar{t}_y$ does not exhibit the property $\pi_e(\mathbf{c}, \tilde{t}_y) = 0$ either. That proves Proposition 2(ii).

Ad (iii). Suppose the tax-only policy $(0, \bar{t}_y)$ is in operation initially, and then an ETS is introduced with the distribution of national caps $\hat{\mathbf{c}}$ (as defined in Proposition 2(i)). We clearly have $A(\hat{\mathbf{c}}, 0) = A(0, \bar{t}_y)$ because $\pi_e(\hat{\mathbf{c}}, \bar{t}_y) = 0$ and there are no permit exports and imports. If another cap distribution $\mathbf{c} \in C$, $\mathbf{c} \neq \hat{\mathbf{c}}$, is chosen, permit exports and imports take place. However, as $\pi_e(\mathbf{c}, \bar{t}_y)$ is zero national incomes remain unchanged. That proves Proposition 2(iii). ■

Proposition 2(i) provides a link between the boxes 1 and 3 of Table 1 through establishing the equivalence of the tax-only policy (which meets the group cap $\bar{c}$) and that particular ETS-only policy $(\hat{\mathbf{c}}, 0)$ under which exports and imports of permits do not take place. That is clearly a very special case which will later be useful for benchmarking, however. Proposition 2(ii) makes use of the tax rate $\bar{t}_y$ of the tax-only policy $(0, \bar{t}_y)$ as well. It answers the question what happens when an

ETS is added to the tax at rate $\bar{t}_y$ and thus provides a link between the boxes 3 and 2 of Table 1 (although the policies $(\mathbf{c}, \hat{t}_y)$ are limiting cases of mixed policies). In that extreme scenario the ETS has no distributional impact at all which gives rise to the conjecture (to be confirmed later) that the distributional potential of varying the assignment of national caps is the smaller the higher is the rate of the overlapping tax.

Under the reasonable assumption of positive income elasticities of both consumer goods Proposition 2(iii) establishes that the tax rate $\bar{t}_y$ of the tax-only policy is the upper bound tax rate for all mixed policies.

Having clarified the relationship of different policies in some extreme scenarios we now turn to more relevant intermediate cases addressing the question as to how equilibria differ when alternative mixed policies are pursued. Primarily we are interested in the distribution of national welfares associated with different intermediate policies. For convenience of notation let us refer to that distribution as $\mathbf{u}(\mathbf{c}, t_y) := [u_1(\mathbf{c}, t_y), \ldots, u_n(\mathbf{c}, t_y)]$, where $u_i(\mathbf{c}, t_y)$ is country i's welfare associated with the equilibrium $E(\mathbf{c}, t_y)$. We seek to answer the question as to what the impact on country i's welfare is of variations

- in the distribution of national emissions caps when the overlapping tax rate remains constant ($\mathrm{d}c_i = -\sum_{j\neq i} \mathrm{d}c_j \neq 0$ and $\mathrm{d}t_y = 0$) and
- in the overlapping tax rate when the distribution of national emissions caps remains constant ($\mathrm{d}c_j = 0$ for all j and $\mathrm{d}t_y \neq 0$).

4.2 Varying National Emissions Caps While Keeping Tax Rates Constant

Consider first policies $(\mathbf{c}, t_y)$ with t_y being fixed. If we start from an equilibrium $E(\mathbf{c}, t_y)$ and consider small changes $\mathrm{d}c_i$ in country i's cap ($i = 1, \ldots, n$) under the constraint $\sum_j \mathrm{d}c_j = 0$, the comparative static effects of $\mathrm{d}c_i$ (Appendix) are[11]

$$\frac{\mathrm{d}u_i}{\lambda_i \mathrm{d}c_i} = \left[\frac{t_y(\alpha_i\delta_i - \beta_i\gamma_i + \alpha_i D_z^i \Delta e_{yi}) + \gamma_i \Delta e_{yi}}{\gamma_i}\right] \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} + \frac{\alpha_i t_y D_z^i \pi_e}{\gamma_i} + \pi_e, \quad (13a)$$

$$\frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} = \frac{\frac{\alpha_i D_z^i \pi_e}{\gamma_i}}{\sum_j \frac{\beta_j\gamma_j - \alpha_j\delta_j - \alpha_j D_z^j \Delta e_{yj}}{\gamma_j}}, \quad (13b)$$

$$\frac{\mathrm{d}p_{xi}}{\mathrm{d}c_i} = \frac{(\delta_i + D_z^i \Delta e_{yi})}{\gamma_i} \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} + \frac{D_z^i \pi_e}{\gamma_i}. \quad (13c)$$

[11] $\alpha_i, \beta_i, \delta_i, \gamma_i$ and λ_i are defined in Appendix and $\Delta e_{yi} := c_i - e_{xi} - e_{yi}$ is the amount of permits exported or imported by country i.

Although the signs of the effects of increasing c_i are ambiguous in general, the terms (13a)–(13c) simplify considerably, if we restrict our attention to quasi-linear utility functions $U^i(x_i, y_i) = V^i(x_i) + y_i$ with V^i being increasing and strictly concave.[12] For this special functional form the income effect of the demand for good X^i is zero ($D_z^i = 0$) which turns Eqs. (13a)–(13c) into

$$\frac{\mathrm{d}u_i}{\lambda_i \mathrm{d}c_i} = \pi_e > 0, \qquad \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} = \frac{\mathrm{d}p_{xi}}{\mathrm{d}c_i} = 0. \tag{13d}$$

The results in (13d) are as expected. Increasing country i's cap increases private income in country i with the straightforward consequence of making its residents better off. In addition, (13d) reveals that quasi-linear utility functions eliminate spillover effects on the market of good X^i.

4.3 Varying Tax Rates in the ETS Sectors While Keeping National Emissions Caps Constant

Suppose next that $\mathbf{c} \in C$ is fixed and that starting from $t_y = 0$, the tax rate t_y is successively raised. Eichner and Pethig (2010) determine the comparative static effects of $\mathrm{d}t_y$ as

$$\frac{\mathrm{d}u_i}{\lambda_i \mathrm{d}t_y} = t_y \left(\frac{\alpha_i \delta_i - \beta_i \gamma_i}{\gamma_i}\right)\left(\frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} + 1\right) + \left(\frac{\alpha_i t_y D_z^i + \gamma_i}{\gamma_i}\right) \Delta e_{yi} \frac{\mathrm{d}\pi_e}{\mathrm{d}t_y}, \tag{14a}$$

$$\frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} = -\frac{1}{1 + \frac{\sum_j \frac{\alpha_j D_z^i}{\gamma_j} \Delta e_{yj}}{\sum_j \frac{\alpha_j \delta_j - \beta_j \gamma_j}{\gamma_j}}}, \tag{14b}$$

$$\frac{\mathrm{d}p_{xi}}{\mathrm{d}t_y} = \frac{\delta_i + \Delta e_{yi} D_z^i}{\gamma_i} \frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} + \frac{\delta_i}{\gamma_i}. \tag{14c}$$

Similar as in the case of Eqs. (13a)–(13c) the sign of the comparative static effects (14a)–(14c) is ambiguous. In a numerical three-country example Eichner and Pethig (2010) identify a country that exports permits initially but eventually imports permits when the tax rate t_y is successively increased. With the rising tax that country's welfare first declines but then increases.

[12] It may be possible to derive from (13a)–(13c) more informative results for functional forms that are less restrictive than quasi-linear utility functions. However, we consider the latter sufficient for the purpose of the present paper because our focus is on distributional *equivalence* of policies rather than on a full characterization of the distributional *impacts* of those policies.

Again, for quasi-linear utility functions there are no interdependence effects on the market of good X^i such that Eqs. (14a)–(14c) simplify to

$$\frac{\mathrm{d}u_i}{\mathrm{d}t_y}\begin{cases} > 0, & \text{if country } i \text{ imports permits,} \\ < 0, & \text{if country } i \text{ exports permits,} \end{cases} \tag{14d}$$

$$\frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} = -1 \qquad \text{and} \qquad \frac{\mathrm{d}p_{xi}}{\mathrm{d}t_y} = 0. \tag{14e}$$

We summarize the preceding results in

Proposition 3 *Every policy* $(\mathbf{c}, t_y) \in C \times [0, \bar{t}_y]$ *has an impact on the welfare distribution* $\mathbf{u}(\mathbf{c}, t_y)$ *via the policy parameter* $\mathbf{c} \in C$ *as well as via the policy parameter* $t_y \in [0, \bar{t}_y]$. *In the case of quasi-linear preferences enlarging a country's emissions cap (at the expense of the other countries' caps) always enhances its welfare, whereas a country gains [loses] from increasing the overlapping tax, if and only if it imports [exports] permits. Under more general assumptions on preferences, market interdependence effects render ambiguous the distributional effect of both policy parameters.*

We have thus demonstrated that changes in the welfare profile $\mathbf{u}(\mathbf{c}, t_y)$ can be brought about either by varying t_y while keeping $\mathbf{c}$ constant or by varying $\mathbf{c}$ while keeping t_y constant (setting perhaps $t_y = 0$). This observation suggests to examine the possibility of neutralizing the welfare effects of an exogenous change in t_y by an appropriate change in $\mathbf{c}$. More generally speaking, in the next section we seek to identify and to characterize subsets of $C \times [0, \bar{t}_y]$ consisting of all mixed policies that exhibit the same welfare profile.

5 Welfare-Neutral (Mixed) Policies

To avoid clumsy wording we call two (mixed) policies $(\mathbf{c}, t_y), (\mathbf{c}', t_y') \in C \times [0, \bar{t}_y]$ welfare neutral, if $\mathbf{u}(\mathbf{c}, t_y) = \mathbf{u}(\mathbf{c}', t_y')$. Obviously, there is a set of welfare-neutral mixed policies for every attainable welfare profile $\mathbf{u} = (u_1, \ldots, u_n)$. To specify these sets, we first consider the following question: Is it possible to start out from a mixed policy $(\mathbf{c}, t_y) \in C \times]0, \bar{t}_y[$ and to find $\tilde{\mathbf{c}} \in C, \tilde{\mathbf{c}} \neq \mathbf{c}$, such that $\mathbf{u}(\tilde{\mathbf{c}}, 0) = \mathbf{u}(\mathbf{c}, t_y)$? We have provided the answer already in Proposition 2(iii) for the extreme policies $(\mathbf{c}, \bar{t}_y)$, $\mathbf{c} \in C$. Now we will show in Proposition 4(i) below the existence of such a mapping for more relevant intermediate policies. That result leaves open the question, however, whether there is more than one mixed policy $(\mathbf{c}, t_y)$ which exhibits the same utility profile as the ETS-only policy $(\tilde{\mathbf{c}}, 0)$. The affirmative answer will be given in Proposition 4(ii). As in Proposition 1 we denote by D^e the set of cost-effective policies $(\mathbf{c}, t_y)$ satisfying $\mathbf{c} \in C$ and $t_y = 0$ and further define D^t to be the set of cost-effective policies satisfying $\mathbf{c} \in C$ and $t_y \in]0, \bar{t}_y]$. Clearly, $D^e \cup D^t$ is the full set of cost-effective policies. Using that notation we establish

Proposition 4

(i) There exists a function $F : D^t \longrightarrow D^e$ such that

$$\mathbf{u}\left[F(\mathbf{c}, t_y)\right] = \mathbf{u}(\mathbf{c}, t_y).$$

(ii) There exists a correspondence $\bar{F} : D^e \longrightarrow D^t$ such that

$$\mathbf{u}(\mathbf{c}, t_y) = \mathbf{u}(\tilde{\mathbf{c}}, 0) \quad \forall\ (\mathbf{c}, t_y) \in \bar{F}(\tilde{\mathbf{c}}, 0) \subset D^t.$$

$\bar{F}(\tilde{\mathbf{c}}, 0)$ has the property that for every $t_y \in]0, t_y(\tilde{\mathbf{c}})[$ there exists $\mathbf{c}' \in C$ such that $(\mathbf{c}', t_y) \in \bar{F}(\tilde{\mathbf{c}}, 0)$. $t_y(\tilde{\mathbf{c}})$ is defined as

$$t_y(\tilde{\mathbf{c}}) := \min_i \frac{\tilde{\pi}_e + \tilde{c}_i}{\tilde{e}_{xi} + \tilde{e}_{yi}} \in]0, \tilde{\pi}_e[,$$

where $\tilde{\pi}_e$, $\tilde{e}_{xi}$ and $\tilde{e}_{yi}$ are the equilibrium values under the ETS-only policy $(\tilde{\mathbf{c}}, 0)$.

Proof

Ad (i). Consider $(\mathbf{c}, t_y) \in D^t$ and define $\tilde{\mathbf{c}}$ by

$$\tilde{c}_i := \frac{\pi_e c_i + t_y(e_{xi} + e_{yi})}{\pi_e + t_y}, \quad i = 1, \ldots, n. \tag{15}$$

Also denote the equilibrium values associated with $(\tilde{\mathbf{c}}, 0)$ by $\tilde{\pi}_e$, $\tilde{p}_{xi}$ etc. Proposition 4(i) follows from

Lemma 1

(a) If $(\mathbf{c}, t_y) \in D^t$ and $\tilde{\mathbf{c}}$ satisfies (15), then $(\tilde{\mathbf{c}}, 0) \in D^e$.
(b) If $(\mathbf{c}, t_y) \in D^t$ and $\tilde{\mathbf{c}}$ satisfies (15), then $E(\tilde{\mathbf{c}}, 0) = E(\mathbf{c}, t_y)$.

Ad (a). Summing (15) over i and invoking (2) and (3) immediately yields

$$\sum_j \tilde{c}_j = \frac{\pi_e \sum_j c_j + t_y \sum_j \left(e_{xj} + e_{yj}\right)}{\pi_e + t_y} = \frac{\left(\pi_e + t_y\right) \sum_j c_j}{\pi_e + t_y} = \bar{c}.$$

Ad (b). Consider $\tilde{\mathbf{c}}$ as defined in (15) and observe that prices satisfy

$$\tilde{\pi}_e = \pi_e + t_y \quad \text{and} \quad \tilde{p}_{xi} = p_{xi} \qquad i = 1, \ldots, n. \tag{16}$$

In the sequel we compare the equilibrium allocations $A(\tilde{\mathbf{c}}, 0)$ and $A(\mathbf{c}, t_y)$. In equilibrium the first-order conditions of profit maximization under policies $(\mathbf{c}, t_y)$ and $(\tilde{\mathbf{c}}, 0)$, respectively, are

$$p_y Y^i\left(e_{yi}\right) = p_e + \pi_e + t_y \quad \text{and} \quad p_{xi} X^i\left(e_{xi}\right) = p_e + t_x = p_e + \pi_e + t_y, \tag{17}$$

$$p_y Y^i(\tilde{e}_{yi}) = p_e + \tilde{\pi}_e \quad \text{and} \quad \tilde{p}_{xi} \tilde{X}^i(\tilde{e}_{xi}) = p_e + \tilde{\pi}_e. \tag{18}$$

Combining (16)–(18) immediately yields

$$\tilde{e}_{hi} = e_{hi} \quad \text{and} \quad \tilde{h}_{si} = h_{si} \quad \text{for} \quad h = x, y. \tag{19}$$

Next, we wish to show that $\tilde{h}_i = h_i(\cdot)$ for $h = x, y$. To that end invoke (16) to transform (15) as follows:

$$\begin{aligned}
&\tilde{\pi}_e \tilde{c}_i = \pi_e c_i + t_y(e_{xi} + e_{yi}) \\
&\quad \Longleftrightarrow \tilde{\pi}_e \tilde{c}_i - \tilde{\pi}_e(e_{xi} + e_{yi}) = \pi_e c_i + t_y(e_{xi} + e_{yi}) - \tilde{\pi}_e(e_{xi} + e_{yi}), \\
&\qquad \tilde{\pi}_e\left(\tilde{c}_i - e_{xi} - e_{yi}\right) = \pi_e(c_i - e_{xi} + e_{yi}).
\end{aligned} \tag{20}$$

If policy $(\mathbf{c}, t_y)$ is given, the consumer's income is

$$z_i = p_{xi} x_{si} + p_y y_{si} - p_e(e_{xi} + e_{yi}) - \pi_e(c_i - e_{xi} - e_{yi}). \tag{21}$$

If policy $(\tilde{\mathbf{c}}, 0)$ is given, the income is

$$\tilde{z}_i = \tilde{p}_{xi} \tilde{x}_{si} + p_y \tilde{y}_{si} - p_e(\tilde{e}_{xi} + \tilde{e}_{yi}) - \tilde{\pi}_e(\tilde{c}_i - \tilde{e}_{xi} - \tilde{e}_{yi}).$$

From (16), (19), (20) and (21) follows $\tilde{z}_i = z_i$. Consequently, the consumer's budget constraint under policy $(\mathbf{c}, t_y)$, is the same as under policy $(\tilde{\mathbf{c}}, 0)$ since prices satisfy (16). The straightforward conclusion is $\tilde{x}_i = x_i$ and $\tilde{y}_i = y_i$. Thus we have shown that $A(\tilde{\mathbf{c}}, 0) = A(\mathbf{c}, t_y)$ and $E(\tilde{\mathbf{c}}, 0) = E(\mathbf{c}, t_y)$.

Ad (ii). Consider $(\tilde{\mathbf{c}}, 0) \in D^e$, denote the corresponding equilibrium values by $\tilde{\pi}_e, \tilde{p}_{xi}$, etc. and define a policy $(\mathbf{c}, t_y)$ by

$$c_i = \frac{\tilde{\pi}_e \tilde{c}_i - (\tilde{e}_{xi} + \tilde{e}_{yi}) t_y}{\tilde{\pi}_e - t_y}, \quad i = 1, \ldots, n. \tag{22}$$

We prove the following

Lemma 2

(a) If $(\tilde{\mathbf{c}}, 0) \in D^e$ *and* $(\mathbf{c}, t_y)$ *satisfies (22), then* $(\mathbf{c}, t_y) \in D^t$.
(b) If $(\tilde{\mathbf{c}}, 0) \in D^e$ *and* $(\mathbf{c}, t_y)$ *satisfies (22), then* $E(\mathbf{c}, t_y) = E(\tilde{\mathbf{c}}, 0)$.

Ad (a). Summing (22) over i and invoking (2) and (3) proves Lemma 2(a).

Ad (b). First we infer from (22) that there is some country i for which $c_i < 0$, if and only if $t_y \geq t_y(\tilde{\mathbf{c}})$. Hence $t_y \in]0, t_y(\tilde{\mathbf{c}})]$ is a necessary equilibrium condition. Next we determine the allocation $(\mathbf{x}_s, \mathbf{e}_x, \mathbf{y}_s, \mathbf{e}_y, \mathbf{x}, \mathbf{y})$ if

$$\pi_e = \tilde{\pi}_e - t_y, \quad p_{xi} = \tilde{p}_{xi} \quad \text{and} \quad t_{xi} = \pi_e + t_y \tag{23}$$

and $t_y \leq t_y(\tilde{\mathbf{c}})$ is given. It is easy to show that all profit-maximizing inputs and outputs are the same: $\mathbf{x}_s = \tilde{\mathbf{x}}_s$, $\mathbf{e}_x = \tilde{\mathbf{e}}_x$, $\mathbf{y}_s = \tilde{\mathbf{y}}_s$, and $\mathbf{e}_y = \tilde{\mathbf{e}}_y$. Since consumer prices are the same, $p_y \equiv 1$ and $\mathbf{p}_x = \tilde{\mathbf{p}}_x$, we conclude that $\mathbf{x} = \tilde{\mathbf{x}}$ and $\mathbf{y} = \tilde{\mathbf{y}}$, iff the consumer incomes are the same, $\mathbf{z} = \tilde{\mathbf{z}}$, i.e. iff for all i

$$\begin{aligned} & p_{xi}x_{si} + y_{si} - p_e(e_{xi} + e_{yi}) - \pi_e(c_i - e_{xi} - e_{yi}) \\ & \quad = \tilde{p}_{xi}\tilde{x}_{si} + \tilde{y}_{si} - p_e(\tilde{e}_{xi} + \tilde{e}_{yi}) - \tilde{\pi}_e(\tilde{c}_i - \tilde{e}_{xi} - \tilde{e}_{yi}). \end{aligned}$$

Since all inputs and outputs are the same, this equation simplifies to

$$\pi_e(c_i - e_{xi} - e_{yi}) = \tilde{\pi}_e(\tilde{c}_i - \tilde{e}_{xi} - \tilde{e}_{yi}),$$

which can readily be turned into (22) after making use of $\pi_e = \tilde{\pi}_e - t_y$ from (23). This observation completes the proof of $E(\mathbf{c}, t_y) = E(\tilde{\mathbf{c}}, 0)$ and Lemma 2. Since Lemma 2 holds for every $t_y \in]0, t_y(\tilde{\mathbf{c}})[$ the second part of Proposition 4(ii) is also proved. ■

According to Proposition 4(i), for each mixed policy $(\mathbf{c}, t_y) \in D^t$ there exists a unique ETS-only policy, $(\tilde{\mathbf{c}}, 0) \in D^e$, yielding the same welfare distribution as the mixed policy. The governments of all countries are indifferent with respect to these policies because each policy produces the same welfare as well as the same resource allocation: $A(\tilde{\mathbf{c}}, 0) = A(\mathbf{c}, t_y)$. Hence switching policies leaves all countries' welfare positions unchanged. In view of $e_{hi}(\mathbf{c}, t_y) = e_{hi}(\tilde{\mathbf{c}}, 0)$ for $h = x, y$ and $i = 1, \ldots, n$, Eq. (15) implies, in fact, that the values of permits imported or exported are the same under the policies $(\mathbf{c}, t_y)$ and $(\tilde{\mathbf{c}}, 0)$. As a consequence, country i's income remains unchanged which leaves the representative consumer's demand for consumption goods unaffected when $p_{xi}(\tilde{\mathbf{c}}, 0) = p_{xi}(\mathbf{c}, t_y)$.

Proposition 4(ii) provides, first of all, the information that we can turn Proposition 4(i) around in the sense that if one starts from the ETS-only policy $F(\mathbf{c}, t_y)$ and chooses the mixed policy $(\mathbf{c}, t_y) \in D^t$, the associated equilibrium allocations remain unchanged. More importantly, for every policy $(\mathbf{c}, 0) \in D^e$ there is a large set $\bar{F}(\mathbf{c}, 0)$ of mixed policies that are welfare-neutral with respect to the ETS-only policy.[13]

[13]Proposition 4(ii) also covers the special case $t = \bar{t}_y$, since $\tilde{c}_i(\bar{t}_y)$ follows immediately from (15) for $\pi_e = 0$ and thus reproduces the result we have already established in Proposition 1 (with much less effort). The important extension of Proposition 2 is that for every policy $(\mathbf{c}, t_y)$ in box 2 of Table 1 there is some policy $(\mathbf{c}', 0)$ in box 1 of Table 1 such that $\mathbf{u}(\mathbf{c}', 0) = \mathbf{u}(\mathbf{c}, t_y)$.

Given the mapping F in Proposition 4(i) from $(\mathbf{c}, t_y) \in D^t$ to $[\tilde{\mathbf{c}}(\mathbf{c}, t_y), 0] \in D^e$, it is natural to ask what the sign and the magnitude are of the differences $c_i - \tilde{c}_i(\mathbf{c}, t_y)$ and $c_{yi} - \tilde{c}_{yi}(\mathbf{c}, t_y)$ and how these differences vary with c_i and t_y, respectively. The answers are provided in

Proposition 5 *Suppose the policy* $(\mathbf{c}, t_y) \in D^t$ *is applied.*

(i) $c_i < \tilde{c}_i(\mathbf{c}, t_y)$ *[*$c_i > \tilde{c}_i(\mathbf{c}_i, t_y)$*], if country i imports [exports] permits.*
(ii) Consider an economy with quasi-linear utility functions. The impact of changes in c_i *and* t_y *on the differences* $c_i - \tilde{c}_i(\mathbf{c}, t_y)$ *and* $c_{yi} - \tilde{c}_{yi}(\mathbf{c}, t_y)$ *are, respectively,*[14]

$$\frac{\mathrm{d}[c_i - \tilde{c}_i(\mathbf{c}, t_y)]}{\mathrm{d}c_i} = \frac{\mathrm{d}[c_{yi} - \tilde{c}_{yi}(\mathbf{c}, t_y)]}{\mathrm{d}c_i} = \frac{t_y}{(\pi_e + t_y)^2}, \tag{24a}$$

$$\frac{\mathrm{d}[c_i - \tilde{c}_i(\mathbf{c}, t_y)]}{\mathrm{d}t_y} = \frac{\mathrm{d}[c_{yi} - \tilde{c}_{yi}(\mathbf{c}, t_y)]}{\mathrm{d}t_y} = \frac{\Delta e_{yi}}{\pi_e + t_y}, \tag{24b}$$

where $\Delta e_{yi} := c_i - e_{xi}(\mathbf{c}, t_y) - e_{yi}(\mathbf{c}, t_y)$.

Proof

Ad (i). Proposition 5(i) is straightforward from rewriting (15) as

$$\tilde{c}_i(\mathbf{c}, t_y) = c_i - \frac{t_y \Delta e_{yi}}{\pi_e + t_y}.$$

Ad (ii). Differentiation of $\tilde{c}_i(\mathbf{c}, t_y)$ with respect to c_i and t_y yields, after some rearrangement of terms,

$$\frac{\mathrm{d}\tilde{c}_i}{\mathrm{d}c_i} = 1 - \frac{t_y \frac{\mathrm{d}\Delta e_{yi}}{\mathrm{d}c_i} - t_y \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} \Delta e_{yi}}{(\pi_e + t_y)^2}, \tag{25a}$$

$$\frac{\mathrm{d}\tilde{c}_i}{\mathrm{d}t_y} = -\frac{\pi_e \left(1 - \frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} \cdot \frac{t_y}{\pi_e}\right)}{(\pi_e + t_y)^2} \Delta e_{yi} - \frac{t_y}{\pi_e + t_y} \cdot \frac{\mathrm{d}\Delta e_{yi}}{\mathrm{d}t_y}. \tag{25b}$$

Differentiate $c_i - \tilde{c}_i(\mathbf{c}, t_y)$ with respect to c_i and t_y, respectively, and make use of (25a) and (25b) to get

$$\frac{\mathrm{d}[c_i - \tilde{c}_i(\mathbf{c}, t_y)]}{\mathrm{d}c_i} = \frac{t_y \frac{\mathrm{d}\Delta e_{yi}}{\mathrm{d}c_i} - t_y \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} \Delta e_{yi}}{(\pi_e + t_y)^2}, \tag{26a}$$

$$\frac{\mathrm{d}[c_i - \tilde{c}_i(\mathbf{c}, t_y)]}{\mathrm{d}t_y} = \frac{\pi_e \left(1 - \frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} \cdot \frac{t_y}{\pi_e}\right)}{(\pi_e + t_y)^2} \Delta e_{yi} + \frac{t_y}{\pi_e + t_y} \frac{\mathrm{d}\Delta e_{yi}}{\mathrm{d}t_y}. \tag{26b}$$

[14]The variation $\mathrm{d}c_i$ is carried out under the constraint $\sum_j \mathrm{d}c_j = 0$.

Consider first (26a). From (13a) and (13c) we infer that $\frac{\mathrm{d}p_{xi}}{\mathrm{d}c_i} = \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} = 0$ if utility functions are quasi-linear $\left(D_z^i = 0\right)$. In addition, (54) in Appendix implies $\frac{\mathrm{d}e_{xi}+\mathrm{d}e_{yi}}{\mathrm{d}c_i} = 0$ for $\frac{\mathrm{d}p_{xi}}{\mathrm{d}c_i} = \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} = 0$ and hence $\frac{\mathrm{d}\Delta e_{yi}}{\mathrm{d}c_i} = 1$ follows. Making use of this information in (26a) we get $\frac{\mathrm{d}[c_i - \tilde{c}_i(\mathbf{c},t_y)]}{\mathrm{d}c_i} = \frac{t_y}{(\pi_e+t_y)^2}$. Next, we differentiate $c_{yi}(\mathbf{c},t_y) - \tilde{c}_{yi}(\mathbf{c},t_y)$ with respect to c_i to obtain $\frac{\mathrm{d}[c_{yi}-\tilde{c}_{yi}(\mathbf{c},t_y)]}{\mathrm{d}c_i} = \frac{\mathrm{d}c_{yi}}{\mathrm{d}c_i} - \frac{\mathrm{d}\tilde{c}_{yi}}{\mathrm{d}c_i}$. Accounting for $c_{yi} = c_i - e_{xi}$, $\tilde{c}_{yi} = \tilde{c}_i - \tilde{e}_{xi}$ and $\frac{\mathrm{d}e_{xi}}{\mathrm{d}c_i} = \frac{\mathrm{d}\tilde{e}_{xi}}{\mathrm{d}c_i} = 0$ from (53) yields $\frac{\mathrm{d}[c_{yi}-\tilde{c}_{yi}(\mathbf{c},t_y)]}{\mathrm{d}c_i} = \frac{\mathrm{d}[c_i-\tilde{c}_i(\mathbf{c},t_y)]}{\mathrm{d}c_i}$ which in turn establishes (24a).

Now we turn to (26b). Implicit in (17) the demand for fossil fuel (and for emissions permits) of country i is given by the functions $E^{xi}(p_{xi}, \pi_e + t_y)$ and $E^{yi}(\pi_e + t_y)$. Totally differentiating these functions E^{xi} and E^{yi} with respect to t_y gives us

$$\frac{\mathrm{d}E^{xi}(p_{xi}, \pi_e + t_y)}{\mathrm{d}t_y} = \frac{\frac{\mathrm{d}\pi_e}{dt_y} + 1}{p_{xi}X_{ee}^i} - \frac{x_{si}}{p_{xi}X_{ee}^i} \cdot \frac{\mathrm{d}p_{xi}}{\mathrm{d}t_y} \quad \text{and}$$
$$\frac{\mathrm{d}E^{yi}(\pi_e + t_y)}{\mathrm{d}t_y} = \frac{\frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} + 1}{Y_{ee}^i}. \tag{27}$$

Obviously, in view of (14e), i.e. $\frac{\mathrm{d}p_{xi}}{\mathrm{d}t_y} = 0$ and $\frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} = -1$ for $i = 1, \ldots, n$, Eqs. (27) imply

$$\frac{\mathrm{d}E^{xi}(p_{xi}, \pi_e + t_y)}{\mathrm{d}t_y} = \frac{\mathrm{d}E^{yi}(\pi_e + t_y)}{\mathrm{d}t_y} = \frac{\mathrm{d}(e_{xi} + e_{yi})}{\mathrm{d}t_y} = 0. \tag{28}$$

Next, we differentiate $\Delta e_{yi} = c_i - e_{xi}(\mathbf{c}, t_y) - e_{yi}(\mathbf{c}, t_y)$ with respect to t_y to obtain $\frac{\mathrm{d}\Delta e_{yi}}{\mathrm{d}t_y} = 0$. Using this information and $\frac{\mathrm{d}\pi_e}{\mathrm{d}t_y} = -1$ in (26b) establishes $\frac{\mathrm{d}[c_i-\tilde{c}_i(\mathbf{c},t_y)]}{\mathrm{d}t_y} = \frac{\Delta e_{yi}}{\pi_e+t_y}$. Invoking the same arguments as above straightforwardly leads to $\frac{\mathrm{d}[c_{yi}-\tilde{c}_{yi}(\mathbf{c},t_y)]}{\mathrm{d}t_y} = \frac{\mathrm{d}[c_i-\tilde{c}_i(\mathbf{c},t_y)]}{\mathrm{d}t_y}$. ■

According to Proposition 5(i), replacing policy $(\mathbf{c}, t_y)$ by the policy $[\tilde{\mathbf{c}}(\mathbf{c}, t_y), 0]$ implies that country i's emissions cap under the new policy $[\tilde{\mathbf{c}}(\mathbf{c}, t_y), 0]$ is greater [smaller] than under the old policy $(\mathbf{c}, t_y)$, if country i imports [exports] permits under the old policy. Moreover, the gap $\left|c_i - \tilde{c}_{yi}(\mathbf{c}, t_y)\right|$ is greater for a permit-exporting country i and smaller for a permit-importing country i, the greater is country i's initial emissions cap, c_i. However, raising the tax rate t_y in the initial policy $(\mathbf{c}, t_y)$ widens the gap $\left|c_i - \tilde{c}_{yi}(\mathbf{c}, t_y)\right|$ for both permit-exporting and permit-importing countries.

It should be (re)emphasized, however, that if the assumption of quasi-linear utility functions is relaxed, the distributional impact of changes in c_i and t_y will be far less clear-cut. Since markets are interdependent, an exogenous change in t_y must be expected to trigger repercussions in other markets so that the crucial presupposition of Proposition 5(ii), $(\mathrm{d}p_{xi}/\mathrm{d}c_i) = (\mathrm{d}p_{xi}/\mathrm{d}t_y) = 0$ *for all i*, is not satisfied, in general. However, if interdependence effects are present, general information cannot be gained from (25a) and (25b) neither on the sign nor on the magnitude of the differential quotients. In particular, the results (24a) and (24b) that changes in $\mathrm{d}c_i$ and $\mathrm{d}t_y$ fully translate into a change in the permit cap $\tilde{c}_{yi}$ must be considered special cases.

To highlight the relevance of Proposition 5 regarding the distributional impact of fixing $\mathbf{c}$ and t_y in policies $(\mathbf{c}, t_y)$, suppose the group of countries has agreed on some distribution of emissions caps, $\mathbf{c} \in C$, satisfying certain equity criteria as in case of the EU burden sharing agreement. If the countries should have determined their "fair" distribution $\mathbf{c}$ without accounting for the preexisting overlapping tax(es), the true distributional impact of the policy $(\mathbf{c}, t_y)$ is unfair according to the equity criteria chosen.

Figure 1 illustrates for the case of two countries the insights of the Propositions 2–5. The line segment $0_1 0_2$ is equal to $\bar{c}$ such that each point on $0_1 0_2$ (e.g. B or E) represents a partition $\mathbf{c} = (c_1, c_2)$ of $\bar{c}$. Furthermore, we set $0_1 G = 0_2 H = \bar{t}_y$ and hence each point in the box $0_1 0_2 HG$ corresponds to some policy $(\mathbf{c}, t_y)$. Moreover, the horizontal straight line GH accounts for the Propositions 2(ii) and 2(iii). Each curve in Fig. 1, such as BD, represents a set of welfare-neutral policies. The ratio u_1/u_2 increases when moving from left to right (e.g. when moving on $0_1 0_2$ toward 0_2 or on CK toward K). By construction of Fig. 1, the point E on $0_1 0_2$ is special,

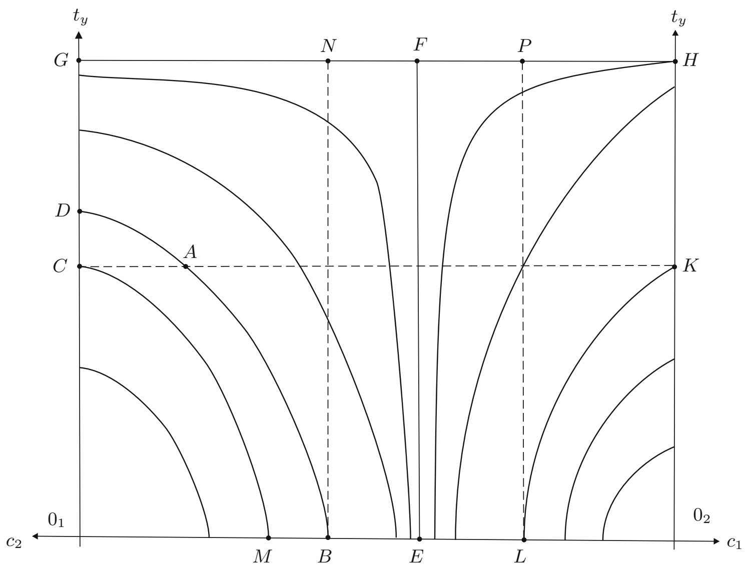


Fig. 1 Welfare implications of cost-effective mixed policies

because in the associated equilibrium no permit exports and imports take place.[15] Country 1 imports permits, iff $c_1 \in [0_1, E[$ and it exports permits, iff $c_1 \in]E, 0_2]$. Located on the vertical straight line from E to F are all mixed policies the associated equilibriums of which have in common the same welfare profile and the absence of permit exports and imports.[16] It follows that the equilibrium allocations and welfare distributions are the same along the lines EF and GH. Note also that according to Proposition 5(i) all welfare indifference curves to the left [right] of EF are downward [upward] sloping.[17]

To illustrate Proposition 4(i) in Fig. 1 suppose that the point A represents the initial mixed policy $(\mathbf{c}, t_y) \in D^t$. Then $F(\mathbf{c}, t_y)$ corresponds to the point B. It is clear that one can also move backward from B to A. But more importantly, equivalent to the ETS-only policy B are all mixed policies that correspond to a point on the curve BD. Hence if we associate the policy $(\tilde{\mathbf{c}}, 0)$ with the point B, the set $\bar{F}(\tilde{\mathbf{c}}, 0)$ corresponds to the set of all points on the curve BD.

Figure 1 can also be used to illustrate the range of possible welfare (re)distributions following from variations in t_y for constant $\mathbf{c}$, or in $\mathbf{c}$ for constant t_y. Obviously, if $t_y = 0$, one can attain any feasible welfare ratio u_1/u_2 by the choice of $\mathbf{c} \in C$ (i.e. by moving along $0_1 0_2$). However, the greater is the tax rate t_y, the smaller becomes the range of u_1/u_2 that can be attained by changes in $\mathbf{c} \in C$. For example, if $t_y = 0_1 C$ in Fig. 1, the welfare distributions that can be generated by moving from C to K are those which are generated by ETS-only policies when moving from M to L. If $t_y = 0_1 G = \bar{t}_y$, variations of $\mathbf{c} \in C$ do not change the welfare distribution at all. Consider finally the scenario in which $\mathbf{c} \in C$ is kept constant and t_y is successively raised from $t_y = 0$ to $\bar{t}_y$. In Fig. 1 that scenario would correspond to starting at some initial ETS-only policy, such as B or L, and moving toward N and P, respectively. Clearly, the move from B to N [L to P] implies a change in welfare distribution equal to that which is implied by the move from B to E [L to E]. As an implication, if we start at point B and raise the tax rate t_y successively, the permit importing country 1 gains, while the permit exporting country loses. The welfares change in opposite directions, if the point of departure lies to the right of E, e.g. L in Fig. 1. Moreover, the smaller is the volume of permit exports and imports in some given initial ETS-only policy, the smaller is the potential of the overlapping emissions tax to bring about changes in the welfare distribution.

[15] Point E represents the utility profile $\mathbf{u}(\hat{\mathbf{c}}, 0)$ with $\hat{\mathbf{c}}$ as specified in Proposition 2(i). Note also that point F represents $\mathbf{u}(\hat{\mathbf{c}}, \bar{t}_y)$ satisfying $\mathbf{u}(\hat{\mathbf{c}}, \bar{t}_y) = \mathbf{u}(\hat{\mathbf{c}}, 0)$ according to Proposition 2(iii).

[16] As implied by Eqs. (14a) and (14b), this is true for general (i.e. not only for quasi-linear) utility functions.

[17] This is why in the two-country case illustrated in Fig. 1 welfares are strictly monotone in both national caps and tax rates even if markets are interdependent.

6 Methods of Measuring the (Re)Distributional Impact of Carbon Emissions Control

This section focuses on cost-effective policies again and aims at *measuring* the distributional impact of those policies $(\mathbf{c}, t_y) \in D^t$. Our finding in Proposition 2(ii) that the equilibrium associated to all policies $(\mathbf{c}, \bar{t}_y)$ is independent of $\mathbf{c}$ suggests taking the distributional impact of the tax-only policy as a benchmark for assigning national emissions caps. Recall that, according to Proposition 2, a given policy $(\mathbf{c}, t_y) \in D^t$ is equivalent to a pure ETS with $\tilde{\mathbf{c}}(\mathbf{c}, t_y) \in D^e$, and that the tax-only policy $\bar{t}_y$ is equivalent to the particular ETS-only policy $[\tilde{\mathbf{c}}(\bar{t}_y), 0]$. From these observations the following measure of distribution is straightforward.

Measure I of Distribution *Taking as a benchmark the policy of implementing $\bar{c}$ with an emissions tax only which is uniform across all sectors and all countries, the redistributional implication of policy $(\mathbf{c}, t_y) \in D^t$ is measured by*

$$\left[\tilde{\mathbf{c}}(\mathbf{c}, t_y) - \tilde{\mathbf{c}}(\bar{t}_y)\right] \in \mathbb{R}^n. \tag{29}$$

Under conditions specified in Sect. 3 we know that switching from the tax-only policy to the policy $(\mathbf{c}, t_y) \in D^t$ makes country i better [worse] off, if and only if $\tilde{c}_i(\mathbf{c}, t_y) > \tilde{c}_i(\bar{t}_y)$ $[\tilde{c}_i(\mathbf{c}, t_y) < \tilde{c}_i(\bar{t}_y)]$. The upside of Measure I is to translate the tax-only policy and mixed policies into shares of permit endowments. Its downside is, however, that its link to the utility distribution is not unambiguous under general forms of utility functions and that it is not a monetary measure.

These limitations are overcome by another straightforward measure that also takes as its benchmark the welfare associated with the tax-only policy. To construct that measure we first introduce a vector of transfer payments $\boldsymbol{\theta} := (\theta_{y1}, \ldots, \theta_{yn}) \in \mathbb{R}^n$ in an equilibrium with policy $(\mathbf{c}, t_y)$. As a result, the welfare of country i becomes equal to

$$u_i(\mathbf{c}, t_y; \theta_i) := U^i\left[D^i(\cdot), z_i(\mathbf{c}, t_y) + \theta_i - p_{xi}(\mathbf{c}, t_y)D^i(\cdot)\right], \tag{30}$$

when it receives the positive or negative transfer θ_i. In (30) $D^i(\cdot) := D^i\left[p_{xi}(\mathbf{c}, t_y), z_i(\mathbf{c}, t_y) + \theta_i\right]$ is the demand for good X^i and $z_i(\mathbf{c}, t_y) + \theta_i - p_{xi}(\mathbf{c}, t_y)D^i(\cdot)$ is the demand for good Y^i.

Measure II of Distribution *Taking as a benchmark the policy of implementing $\bar{c}$ with an emissions tax only which is uniform across all sectors and all countries, the redistributional implication of policy $(\mathbf{c}, t_y) \in D^t$ is measured by the monetary transfer $\boldsymbol{\theta}(\mathbf{c}, t_y) := [\theta_1(\mathbf{c}, t_y), \ldots, \theta_n(\mathbf{c}, t_y)] \in \mathbb{R}^n$, where for all i the monetary* *transfer* $\theta_i = \theta_i(\mathbf{c}, t_y)$ *is defined by*[18]

$$u_i[\tilde{\mathbf{c}}(\bar{t}_y), 0; \theta_i] = u_i(\mathbf{c}, t_y; 0). \tag{31}$$

According to (31) $\theta_i(\mathbf{c}, t_y)$ is the amount of money country i needs to receive or needs to pay in order to shift its utility from the level $u_i\left[\tilde{\mathbf{c}}(\bar{t}_y), 0; \theta_i\right]$ to the level $u_i(\mathbf{c}, t_y; 0) = u_i(\mathbf{c}, t_y)$ which it actually enjoys in the equilibrium attained under the policy $(\mathbf{c}, t_y)$. Switching from $\left[\tilde{\mathbf{c}}(\bar{t}_y), 0\right]$ to $(\mathbf{c}, t_y)$ creates winners and losers. If $\theta_i(\mathbf{c}, t_y) > 0$, country i loses through that policy switch because it needs the compensation $\theta_i(\mathbf{c}, t_y) > 0$ in order to be indifferent between both policy schemes. Conversely, if $\theta_i(\mathbf{c}, t_y) < 0$, it gains through that policy switch because its income under policy $(\mathbf{c}, t_y)$ needs to be reduced by $\theta_i(\mathbf{c}, t_y)$ to make its utility level under policy $(\mathbf{c}, t_y)$ equal to the level it enjoys under the tax-only policy.

7 Concluding Remarks

Burden sharing is well known to be a crucial precondition for successful international carbon emissions control within the EU as well as world wide. In the present paper we do not address fairness in burden sharing but focus, instead, on the questions preceding the fairness issue, namely what the true national burdens are in hybrid emissions control policies and how to measure them. We show that when an ETS covering only part of all participating countries' economies is combined with an overlapping emissions tax, the net impact on national welfares results from an integrated account of the partial welfare effects of both instruments. Our welfare-neutrality result allows expressing each country's net burden carried in a mixed policy as the net burden it carries in a hypothetical but welfare-neutral ETS-only policy. The distributional impact of a uniform overlapping tax is thus 'translated' into changes in national emissions caps. The national net burdens are shown to be measurable as deviations from the burdens implied by the tax-only policy.

Our paper provides a message for parties involved in negotiations about an agreement on the distribution of national emissions caps in the context of a joint EU-type ETS. When major emissions taxes overlapping with the ETS exist, the negotiated national emissions caps are distorted indicators of national burdens, unless the burdens implicit in the overlapping taxes are properly taken into account. Rational burden sharing negotiations need to consider the 'burden impact' of both instruments. There are reasons to doubt whether the parties in the EU burden sharing

[18] In spirit, θ_i is analogous to the Hicksean equivalent variation.

agreement had at their disposal and/or used all the information about the incidence of their agreed-upon national caps needed to share the burden according to their own fairness criteria. The policy message of the paper is that the parties have to calculate their 'true' net burdens invoking the welfare-neutrality result and the associated measures established above.

In the major part of the paper we assume cost-effective mixed policies to avoid blurring distributional and efficiency effects. Yet as we have noted above, the actual hybrid EU policy is not cost effective because, among other things, the extant national overlapping taxes are not uniform across countries. We were able to show that our procedure of specifying burdens for cost-effective mixed policies can be extended to the empirically relevant scenario of non-uniform taxes by decomposing the total welfare effect into an effect capturing the allocative inefficiency and into another effect that isolates the impact on the distribution of welfares. In the case of cost-*in*effective mixed policies distributional equivalence is combined with an overall efficiency loss that may be distributed by reducing the welfare of all countries at a uniform rate. The economist's recommendation would be, of course, to eliminate the inefficiency through tax harmonization in the first place.

Appendix: The Comparative Statics of Changing the Permit Cap c_i

The cost-effective competitive equilibrium is determined by the equations

$$\sum_j c_j = \sum_j (e_{xj} + e_{yj}), \tag{32}$$

$$x_{si} = x_i \qquad i = 1, \ldots, n, \tag{33}$$

$$x_{si} = X^i(e_{xi}), \qquad i = 1, \ldots, n, \tag{34}$$

$$x_i = D^i(p_{xi}, z_i), \qquad i = 1, \ldots, n, \tag{35}$$

$$z_i = p_{xi}x_{si} + y_{si} - p_e(e_{xi} + e_{yi}) + \pi_e(c_i - e_{xi} - e_{yi}), \quad i = 1, \ldots, n, \tag{36}$$

$$y_{si} = Y^i(e_{yi}), \qquad i = 1, \ldots, n, \tag{37}$$

$$z_i = p_{xi}x_i + y_i, \qquad i = 1, \ldots, n, \tag{38}$$

$$p_{xi}X_e^i(e_{xi}) = p_e + t_x, \qquad i = 1, \ldots, n, \tag{39}$$

$$Y_e^i = p_e + \pi_e + t_y, \qquad i = 1, \ldots, n, \tag{40}$$

$$t_x = \pi_e + t_y. \tag{41}$$

In (32)–(41) good Y is chosen as numeraire. The demand function (35) follows from the first-order condition for utility maximization. It is convenient to compress the system of Eqs. (32)–(41) as follows

$$\sum_j c_j = \sum_j (e_{xj} + e_{yj}), \tag{42}$$

$$X^i(e_{xi}) = D^i(p_{xi}, z_i), \tag{43}$$

$$z_i = p_{xi}X^i(e_{xi}) + Y^i(e_{yi}) - p_e(e_{xi} + e_{yi}) + \pi_e \Delta e_{yi}, \tag{44}$$

$$p_{xi}X^i_e(e_{xi}) = Y^i_e(e_{yi}), \tag{45}$$

$$Y^i_e(e_{yi}) = p_e + \pi_e + t_y, \tag{46}$$

$$y_i = Y^i(e_{yi}) - p_e(e_{xi} + e_{yi}) + \pi_e \Delta e_{yi}, \tag{47}$$

where $\Delta e_{yi} := c_i - e_{xi} - e_{yi}$. Our aim is to determine through a comparative static analysis the impact of exogenous variations in the caps c_i subject to the constraint $\sum_j \mathrm{d}c_j = 0$. To that end (42)–(46) are totally differentiated.

$$\sum_j (\mathrm{d}e_{xj} + \mathrm{d}e_{yj}) = 0, \tag{48}$$

$$X^i_e \mathrm{d}e_{xi} - D^i_p \mathrm{d}p_{xi} - D^i_z \mathrm{d}z_i = 0, \tag{49}$$

$$\mathrm{d}z_i - x_i \mathrm{d}p_{xi} - t_y(\mathrm{d}e_{xi} + \mathrm{d}e_{yi}) - \Delta e_{yi}\mathrm{d}\pi_e - \pi_e \mathrm{d}c_i = 0, \tag{50}$$

$$X^i_e \mathrm{d}p_{xi} + p_{xi}X^i_{ee}\mathrm{d}e_{xi} - Y^i_{ee}\mathrm{d}e_{yi} = 0, \tag{51}$$

$$Y^i_{ee}\mathrm{d}e_{yi} - \mathrm{d}\pi_e = 0. \tag{52}$$

Inserting $\mathrm{d}e_{yi} = \frac{\mathrm{d}\pi_e}{Y^i_{ee}}$ from (52) in (51) yields

$$\mathrm{d}e_{xi} = \frac{\mathrm{d}\pi_e}{p_{xi}X^i_{ee}} - \frac{X^i_e \mathrm{d}p_{xi}}{p_{xi}X^i_{ee}}. \tag{53}$$

Summation of $\mathrm{d}e_{xi}$ from (53) and $\mathrm{d}e_{yi}$ from (52) gives

$$\mathrm{d}e_{xi} + \mathrm{d}e_{yi} = \alpha_i \mathrm{d}p_{xi} - \beta_i \mathrm{d}\pi_e, \tag{54}$$

where $\alpha_i := -\frac{X^i_e}{p_{xi}X^i_{ee}} > 0$ and $\beta_i := -\left(\frac{1}{Y^i_{ee}} + \frac{1}{p_{xi}X^i_{ee}}\right) > 0$. Inserting (54) in (48) we obtain

$$\frac{\sum_j \alpha_j \mathrm{d}p_{xj}}{\sum_j \beta_j} = \mathrm{d}\pi_e. \tag{55}$$

Next, we take advantage of (54) to turn (50) into

$$\mathrm{d}z_i = (x_i + \alpha_i t_y)\mathrm{d}p_{xi} + (\Delta e_{yi} - \beta_i t_y)\mathrm{d}\pi_e + \pi_e \mathrm{d}c_i. \tag{56}$$

We make use of (53) and (56) to transform (49) into

$$\mathrm{d}p_{xi} = \frac{(\delta_i + D_z^i \Delta e_{yi})}{\gamma_i}\mathrm{d}\pi_e + \frac{D_z^i \pi_e}{\gamma_i}\mathrm{d}c_i, \tag{57}$$

where $\delta_i := \alpha_i - \beta_i t_y D_z^i$ and $\gamma_i := \alpha_i X_e^i - D_p^i - (x_i + \alpha_i t_y)D_z^i$. We insert (57) into (55) to obtain, after some rearrangement of terms,

$$\mathrm{d}\pi_e \left[\sum_j \left(\frac{\beta_j \gamma_j - \alpha_j \delta_j - \alpha_j D_z^j \Delta e_{yj}}{\gamma_j}\right)\right] = \sum_j \frac{\alpha_j D_z^j \pi_e}{\gamma_j}\mathrm{d}c_j. \tag{58}$$

Next, we differentiate the utility function $u_i = U^i(x_i, y_i)$ and use the first-order condition of the consumer's utility maximization problem to get

$$\frac{\mathrm{d}u_i}{\lambda_i} = p_{xi}\mathrm{d}c_i + \mathrm{d}y_i, \tag{59}$$

where λ_i is the marginal utility of income. From (34), (39) and (41) we infer

$$\mathrm{d}x_i = X_e^i \mathrm{d}e_{xi} = \frac{p_e + \pi_e + t_y}{p_{xi}}\mathrm{d}e_{xi}. \tag{60}$$

From (47) we obtain with the help of (46)

$$\mathrm{d}y_i = t_y \mathrm{d}e_{yi} - (p_e + \pi_e)\mathrm{d}e_{xi} + \Delta e_{yi}\mathrm{d}\pi_e + \pi_e \mathrm{d}c_i. \tag{61}$$

Inserting (61) and (60) in (59) yields after some rearrangement of terms

$$\frac{\mathrm{d}u_i}{\lambda_i \mathrm{d}c_i} = t_y \frac{\mathrm{d}e_{xi} + \mathrm{d}e_{yi}}{\mathrm{d}c_i} + \Delta e_{yi}\frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} + \pi_e. \tag{62}$$

From (54) it follows that

$$\frac{\mathrm{d}e_{xi} + \mathrm{d}e_{yi}}{\mathrm{d}c_i} = \alpha_i \frac{\mathrm{d}p_{xi}}{\mathrm{d}c_i} - \beta_i \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i}. \tag{63}$$

(57) yields

$$\frac{\mathrm{d}p_{xi}}{\mathrm{d}c_i} = \frac{(\delta_i + D_z^i \Delta e_{yi})}{\gamma_i}\frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} + \frac{D_z^i \pi_e}{\gamma_i}. \tag{64}$$

Making use of (64) in (63) yields

$$\frac{\mathrm{d}e_{xi} + \mathrm{d}e_{yi}}{\mathrm{d}c_i} = \left(\frac{\alpha_i \delta_i - \beta_i \gamma_i + \alpha_i D_z^i \Delta e_{yi}}{\gamma_i} \right) \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} + \frac{\alpha_i D_z^i \pi_e}{\gamma_i}. \tag{65}$$

Finally, taking advantage of (65) in (62) establishes

$$\frac{\mathrm{d}u_i}{\lambda_i \mathrm{d}c_i} = \left[\frac{t_y(\alpha_i \delta_i - \beta_i \gamma_i + \alpha_i D_z^i \Delta e_{yi}) + \gamma_i \Delta e_{yi}}{\gamma_i} \right] \frac{\mathrm{d}\pi_e}{\mathrm{d}c_i} + \frac{\alpha_i t_y D_z^i \pi_e}{\gamma_i} + \pi_e. \tag{66}$$

References

Babiker, M., Metcalf, G. E., & Reilly, J. (2003). Tax distortions and global climate policy. *Journal of Environmental Economics and Management, 46*, 269–287.

Bento, A. M., & Jacobsen, M. (2007). Ricardian rents, environmental policy and the 'double-dividend' hypothesis. *Journal of Environmental Economics and Management, 53*, 17–31.

Böhringer, C., Koschel, H., & Moslener, U. (2008). Efficiency losses from overlapping economic instruments in European carbon emissions regulation. *Journal of Regulatory Economics, 33*, 299–317.

Bovenberg, A. L., & de Mooij, R. (1994). Environmental levies and distortionary taxation. *American Economic Review, 84*, 1085–1089.

Caplan, A. J., Cornes, R. C., & Silva, E. C. D. (2003). An ideal Kyoto protocol: emissions trading, redistributive transfers and global participation. *Oxford Economic Papers, 55*, 216–234.

Chichilnisky, G., & Heal, G. (1994). Who should abate carbon emissions? An international viewpoint. *Economics Letters, 44*, 443–449.

Eichner, T., & Pethig, R. (2009). Efficient CO_2 emissions control with emissions taxes and international emissions trading. *European Economic Review, 53*, 625–635.

Eichner, T., & Pethig, R. (2010). EU-type carbon emissions trade and the distributional impact of overlapping emissions taxes. *Journal of Regulatory Economics, 37*, 287–315.

EU. (2003). Directive establishing a scheme for greenhouse gas emission allowances within the Community and amending Council Directive 96/61/EC. Council Directive 2003/87/EC, European Commission, Brussels.

IEA. (2011). Energy efficiency and carbon pricing, information paper. Paris.

Johnstone, N. (2003). The use of tradable permits in combination with other environmental policy instruments. Report ENV/EPOC/WPNEP (2002)28/FINAL, OECD, Paris.

Lehmann, P. (2012). Justifying a policy mix for pollution control: A review of economic literature. *Journal of Economic Surveys, 26*, 71–97.

Mandell, S. (2008). Optimal mix of emissions taxes and cap-and-trade. *Journal of Environmental Economics and Management, 56*, 131–140.

Marklund, P. O., & Samakovlis, E. (2007). What is driving the EU burden-sharing agreement: Efficiency or equity? *Journal of Environmental Economics and Management, 85*, 317–329.

Nordhaus, W. D. (2006). Life after Kyoto: Alternative approaches to global warming policies. *American Economic Review: Papers and Proceedings, 96*, 31–34.

Peterson, S., & Klepper, G. (2007). Distribution matters - Taxes vs. emissions trading in post Kyoto climate regimes. Kiel Working Paper No. 1380, Kiel Institute for the World Economy.

Phylipsen, G. J. M., Bode, J. W., Blok, K., Merkus, H., & Metz, B. (1998). A Triptych sectoral approach to burden differentiation; GHG emissions in the European bubble. *Energy Policy, 26*, 929–943.

Rosendahl, K. E. (2008). Incentives and prices in an emissions trading scheme with updating. *Journal of Environmental Economics and Management, 56*, 69–82.

Shiell, L. (2003). Equity and efficiency in international markets for pollution permits. *Journal of Environmental Economics and Management, 46*, 38–51.

Sorell, S., & Sijm, J. (2003). Carbon trading in the policy mix. *Oxford Review of Economic Policy, 19*, 420–437.

Thinking Local but Acting Global? The Interplay Between Local and Global Internalization of Externalities

Karen Pittel and Dirk Rübbelke

1 Introduction

When discussing policies that aim at reducing fossil fuel combustion, politicians as well as academics focus almost exclusively on the effects of these policies on climate change. In reality however, many activities that reduce emissions of greenhouse gases also reduce local damages from the combustion of fossil fuels. Emissions of nitrogen oxides, for example, harm the climate but also cause acid rain and respiratory problems. In this paper, we analyze the implications of considering both, the global as well as the local effects of burning fossil fuels, when designing environmental policies.

In order to capture the different nature of local and global damages in a unified framework, we employ a so-called impure public good approach. The seminal approach for the analysis of such goods has been developed by Cornes and Sandler (1984). They define a typical consumer's utility function in terms of characteristics instead of marketed commodities as suggested by Lancaster (1966, 1971). The notion of 'characteristics' takes account of the fact that goods regularly possess a number of different properties that are relevant to choice. The resulting differentiation of seemingly homogeneous goods (like eggs) justifies price-differentiation (Gorman 1980). An important group of goods raising economists' interests are impure public goods that comprise private as well as public characteristics. In our

K. Pittel (✉)
Ifo Institute – Leibniz Institute for Economic Research and University of Munich, Poschingerstrasse 5, 81679 Munich, Germany
e-mail: pittel@ifo.de

D. Rübbelke
Chair for Economics Especially Resource Economics, TU Bergakademie Freiberg, Freiberg, Germany
e-mail: Dirk.Ruebbelke@vwl.tu-freiberg.de

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*, DOI 10.1007/978-3-319-49442-5_14

case, local damages are considered to be private characteristics as they can be perfectly internalized by a local or national regulator. Transnational damages, on the other hand, are assumed to be of a public nature for which no institution exists that has the legal authority to regulate the emissions on a global scale.

Despite their complexity, joint production models involving characteristics of divergent degree of publicness, have been applied in various fields in the economics discipline, e.g. Posnett and Sandler (1986) and Andreoni (1989) apply the impure public good approach developed by Cornes and Sandler (1984) to the field of charitable giving. Impure public good models have also been employed to analyze, e.g. climate protection (e.g. Sandler 1996), green markets (Kotchen 2006), military alliances (Sandler and Murdoch 1990; Sandler and Hartley 2001), terrorism (Rübbelke 2005; Pittel and Rübbelke 2006), sustainable development (Rive and Rübbelke 2010) and theater arts (Pugliese and Wagner 2011).

Cornes and Sandler (1994) conduct a comparative statics analysis where the divergence of results is largely caused by different degrees of substitutability/complementarity between the impure public good's private and public characteristics. Extensions of the impure public good model including alternative technologies generating characteristics of the impure public good, have been provided—amongst others—by Vicary (1997), Rübbelke (2003) and Kotchen (2005). Auld and Eden (1990) analyse a world of two impure public goods each of which has three characteristics. In a recent paper, Cornes (2016) expounds the role that the aggregative games approach may play concerning the analysis of impure public goods.

In the present paper, we apply the impure public good approach in a dynamic framework. We set up a model that accounts for local flow pollution as well as for global stock pollution and explicitly considers two types of abatement activities that differ with respect to their implications regarding local and global pollution mitigation. We take climate protection as one prominent example for our analysis.

Using our approach, we characterize policies that consider the returns from abating global and local pollution simultaneously and that result in an optimal abatement mix. We also derive consequences of different degrees of internalization. A national regulator, for example, might focus foremost on reducing local pollution while potentially taking local but not international damages from global pollutants into account. We show how the design of policies depends on the scope of internalization. Thus, we argue that local and global environmental policies should not be treated separately but rather in a unified framework.

In the literature on climate change, additional benefits from climate protection are mostly of a local or regional nature (see, e.g., IPCC 1996; Pearce 1992) and are often referred to as ancillary benefits, implying that the main benefit lies in the reduction of greenhouse gases. A more neutral term is 'co-benefits' (see IPCC 2001) which leaves undecided whether the primary target is the mitigation of global or local pollution.

Although often neglected, these so-called co-benefits are estimated to be of considerable size (see, e.g., Pearce 2000). Decreasing fossil fuel combustion in the transport sector by increasing technological efficiency, for example, not only reduces the emissions of greenhouse gases like CO_2, CH_4 and NO_2 but also reduces emissions of regional or local pollutants like particulate matter, SO_2 and NO_x. As a result, negative effects of these pollutants like health problems, acid rain, and surface corrosion are mitigated as well (see, e.g., Rübbelke 2002). Similarly, afforestation and deforestation not only enhance carbon sequestration but can also reduce soil erosion and foster biodiversity. Consequently, a comprehensive analysis of the costs and benefits of, e.g., the global warming problem should incorporate co-benefits from preventing greenhouse gas emissions (see also Morgenstern 1991; Plambeck et al. 1997).

In our paper we not only include local and global pollution but also two types of abatement. Abatement either affects local pollution only, or local and global pollution simultaneously. The former could, for example, be filters that reduce the emission of particulate matter; the latter could be the aforementioned reduction of fuel combustion. The main target of the latter could be global or local, depending on the aim of the policy maker. By considering these different pollution and abatement types and their interrelations, we can analyze their effects on different internalization strategies and environmental policies.

In order to include not only the intertemporal spillovers from CO_2-accumulation but also their transnational nature, we consider two countries that each produce and pollute. To keep the focus on the internalization of the pollution induced externalities, we employ an AK-type endogenous growth model in which no other market failures arise. We also abstract from any flows of goods or capital between the countries, such that the economies only affect each other through transnational pollution spillovers.

A look at the related literature shows that, so far, most papers that consider both, local as well as global, benefits from pollution abatement have been case studies assessing the level of ancillary benefits for individual regions or countries (e.g., Gielen and Chen 2001; Li 2006) or have been analytical models which employed static approaches neglecting dynamic implications (e.g., Pittel and Rübbelke 2008; Finus and Rübbelke 2013). On the other hand, the strand of analytical literature that deals with the dynamics of economic development, and the growth-pollution nexus specifically, usually considers either flow or stock pollution but does not take potential interrelations into account (e.g., Withagen 1995; Smulders and Gradus 1996; Schou 2000, 2002). Furthermore, it is rarely distinguished between local and global pollution as most approaches assume closed economies. Bahn and Leach (2008) consider secondary effects of climate policy due to the reduction of SO_2 emissions in an overlapping generation model. Their model is, however, not analytical solvable, such that the general impact and transmission channels of these secondary effects are not clearly identifiable.

The remainder of the paper is organized as follows: After the introduction of the model in Sect. 2, we consider four different internalization scenarios in Sect. 3. These scenarios differ (a) in the degree of internalization of the global externality and (b) with respect to the symmetry of internalization in the two countries.

In Sect. 4 we consider environmental taxes as one example for policy options to decentralize the internalization scenarios of the previous section. Section 5 concludes.

2 The Model

Two countries i, $i = h,f$, produce a homogeneous output from capital. The input of capital generates two types of pollution which differ with respect to their range of geographical impact. For simplicity we assume the two countries to be identical with respect to their production technologies as well as preferences. It is assumed that neither capital nor goods are exchanged between the two countries, such that we can fully concentrate on the local and global environmental externalities.

The externality created by the first type of pollution (e.g. current emissions of CO_2), P_G, is of a global nature, i.e. it affects production in both countries. Due to the long period of time it takes for emissions like CO_2 to be absorbed in the atmosphere, we assume that these emissions build up a renewable pollution stock, S, that degenerates at rate a. As both countries generate pollution, the pollution stock dynamics are given by

$$\dot{S} = P_G^h + P_G^f - aS \tag{1}$$

with $\dot{S} = \frac{\mathrm{d}S}{\mathrm{d}t}$. For simplicity we assume that capital, K^i, generates pollution in a constant ratio p_G. Global pollution can be reduced by abatement activities, A_{LG}^i. This type of abatement simultaneously decreases the environmental impacts of local pollution and is therefore indexed LG for **L**ocal and **G**lobal (for simplicity, A_{LG}^i will, however, be referred to as 'global' abatement from here onwards). P_G^i reads

$$P_G^i = p_G \frac{K^i}{A_{LG}^i}, \qquad i = h,f. \tag{2}$$

The second type of pollution, P_L^i, gives rise to a local externality that only affects production in the country in which it is generated. Examples for this type of pollution might be the emission of SO_2 or NO_x that lead to, e.g., acid rain in a limited regional range around the point of emission. With these types of pollution in mind, P_L^i is assumed to give rise to a flow externality.[1] Again we assume pollution to be generated in fixed proportions to the input of capital. The environmental impact of local pollution can not only be reduced by abatement activities A_{LG}^i but also by abatement activities A_L^i that reduce local pollution only. P_L^i as a function of capital

[1]This is of course an approximation which seems, however, justifiable when comparing the degradation rates of, e.g., SO_2 induced pollution to CO_2.

and the two types of abatement then reads

$$P_L^i = p_L \frac{K^i}{(A_L^i)^\alpha (A_{LG}^i)^{1-\alpha}}, \qquad 0 < \alpha < 1 \tag{3}$$

with p_L denoting the pollution intensity of capital.

Output, Y^i, is produced using a linear AK-technology in the tradition of Rebelo (1991). This simple representation of the production process was chosen as it allows us to focus on the implications of the pollution-related market failures only. Compared to a set-up economy in which market failures additionally arise from, e.g., knowledge spill-overs or monopolistic competition as in Pittel and Bretschger (2009) or Grimaud and Rougé (2003), this renders the analysis more straightforward. Both, the global pollution stock and local pollution flow, have negative effects on production:

$$Y^i = K^i (P_L^i)^{-\gamma} S^{-\delta}, \qquad \gamma, \delta > 0, \; \gamma + \delta < 1. \tag{4}$$

It is assumed that despite the negative effects of local and global pollution on productivity, the marginal product of capital net of these effects is still positive ($1 - \gamma - \delta > 0$). Output can be used for consumptive, investive and abatement purposes, such that the equilibrium condition for the capital market reads

$$\dot{K}^i = Y^i - C^i - A_L^i - A_{LG}^i. \tag{5}$$

Finally, households in country i derive utility from consumption C^i. The representative household maximizes discounted lifetime utility:

$$\max_c \int_0^\infty \frac{C^i(t)^{1-\sigma}}{1-\sigma} e^{-\rho t} \, \mathrm{d}t \qquad \sigma \neq 1 \tag{6}$$

where ρ denotes the discount rate.

3 The Internalization Scenarios

In the following we distinguish between different types of scenarios: First, it is assumed that the regulator in each country i only internalizes the local pollution externality but completely neglects the global externality. Second, the regulators internalize the effects of local and global pollution on the domestic economy but do not take into account that domestic CO_2-emissions also cause damages abroad. Third, we consider a global social planner who internalizes all externalities perfectly. In a final subsection, we assume internalization regimes to differ across countries, i.e. we consider asymmetric internalization.

3.1 Scenario 1: Internalization of the Local Externality Only

The regulator in each country maximizes intertemporal utility of the representative household, (6), subject to the dynamics of the capital stock, (5). After inserting (3) and (4) the corresponding Hamiltonian in Scenario 1 for each country reads

$$H^{i^1}(C^i, K^i, A_L^i, A_{LG}^i, \lambda^i) = \frac{(C^i)^{1-\sigma}}{1-\sigma} e^{-\rho t} \\ + \lambda^i (p_L^{-\gamma} (K^i)^{1-\gamma} (A_L^i)^{\alpha\gamma} (A_{LG}^i)^{(1-\alpha)\gamma} S^{-\delta} - C^i - A_{LG}^i - A_L^i) \tag{7}$$

where λ^i is the shadow price of capital. Optimization gives rise to the following first-order conditions:

$$(C^i)^{-\sigma} e^{-\rho t} = \lambda^i \tag{8}$$

$$\alpha\gamma \frac{Y^i}{A_L^i} = \gamma(1-\alpha)\frac{Y^i}{A_{LG}^i} = 1 \tag{9}$$

$$g_{\lambda^i} = -(1-\gamma)\frac{Y^i}{K^i} \tag{10}$$

and the transversality condition for the capital stock, $\lim_{t\to\infty} \lambda^i K^i = 0$. $g_{\lambda^i} = \frac{\dot{\lambda}^i}{\lambda^i}$ is the growth rate of λ^i.

From (8) and (10) we get the growth rate of consumption

$$g_C^i = \frac{1}{\sigma}\left((1-\gamma)Y_K^i - \rho\right) \tag{11}$$

where $Y_K^i = \frac{Y^i}{K^i}$ denotes the output-capital-ratio. $(1-\gamma)Y_K^i$ gives the marginal product of capital net of local pollution effects.

In Scenario 1 an increase in A_{LG}^i always implies a proportional increase in A_L^i as (9) shows that the two abatement activities will be employed in a constant ratio that is determined by their respective productivity in reducing local pollution:

$$\frac{A_{LG}^i}{A_L^i} = \frac{1-\alpha}{\alpha}. \tag{12}$$

The simplicity of (12) is due to the neglect of stock pollution by the regulator. The mitigating effect of A_{LG}^i on global pollution is thus not reflected in (12).

Using (4) and (12), the ratio of capital and global abatement, $K_A^i = \frac{K^i}{A_{LG}^i}$, can be expressed as a function of the global pollution stock S only:

$$K_A^i = [\gamma(1-\alpha)]^{-\frac{1-\alpha\gamma}{1-\gamma}} \left[\frac{p_L^{\gamma} S^{\delta}}{(\alpha\gamma)^{\alpha\gamma}}\right]^{\frac{1}{1-\gamma}}. \tag{13}$$

As K_A^i only depends on the pollution stock and no country specific variables, the capital-abatement ratio will be identical across countries, i.e. $K_A^h = K_A^f$, at any point in time. Equation (13) also shows the positive relation between K_A^i and the pollution stock, i.e. a higher S goes along with less abatement relative to capital accumulation.

From (9), the capital-abatement ratio can be expressed as $Y_K^i = [\gamma(1-\alpha)K_A^i]^{-1}$ where K_A^i is determined by (13). The dynamics of consumption, (11), can thus be rewritten in terms of the capital-abatement ratio:

$$g_C^i = \frac{1}{\sigma}\left(\frac{1-\gamma}{\gamma(1-\alpha)}\frac{1}{K_A^i} - \rho\right). \tag{14}$$

As K_A^i depends positively on the pollution stock, a higher S implies lower growth.

Along any balanced growth path (BGP), also referred to as long-run equilibrium, all variables grow at constant rates such that $g_Y^i = g_C^i = g_K^i$ and $g_S = 0$ has to hold. The latter condition immediately implies that $g_K^i = g_{A_L}^i = g_{A_{LG}}^i$ along the BGP. Due to the symmetry assumptions, growth rates are also equal across countries. The pollution stock along the BGP is constant and given by

$$S = \frac{p_G(K_A^h + K_A^f)}{a} = 2\frac{p_G K_A^i}{a}. \tag{15}$$

Consequently, K_A^i in the long-run equilibrium can be rewritten as

$$K_A^i = (\gamma(1-\alpha))^{-\frac{1-\alpha\gamma}{1-\gamma-\delta}}\left(\frac{p_L^\gamma\left(\frac{2p_G}{a}\right)^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{\frac{1}{1-\gamma-\delta}}. \tag{16}$$

Equation (16) reaffirms that abatement, A_{LG}^i, and capital, K^i, grow at the same rate along the BGP, rendering K_A^i as well as global pollution constant in the long-run. Constancy of local pollution follows from (12).

The RHS of (16) depends positively on the elasticity of output with respect to stock pollution, δ, as well as on the pollution intensities of capital, p_L and p_G. For any given capital input and level of abatement, an increase in either of these parameters reduces output and thereby the marginal product of abatement. Thus the regulator finds it optimal to increase the capital-abatement ratio until the marginal product of abatement is again equal to its marginal costs. As a result, interestingly, a stronger effect of pollution on production induces a higher capital-abatement ratio and therefore higher pollution.

The same effect arises with respect to γ, the elasticity of output with respect to flow pollution. Yet, an increase in γ also affects the marginal product of abatement positively as abatement becomes more productive, see (9). Depending on which of the two effects dominates, K_A^i rises or falls with γ. As can be seen from $\frac{\partial K_A^i}{\partial \gamma} = \frac{K_A^i[\log(P_L^i)-1/\gamma]}{1-\gamma-\delta}$, the higher the equilibrium pollution flow, the more likely K_A^i increases with γ.

It can be shown that the BGP is locally saddle-path stable by rewriting the dynamic system in terms of S and the consumption-capital ratio, C_K^i, which is constant along the BGP. From (1), (5) and (14) we get for each country (recall that $K_A^h = K_A^f$ and thus $C_K^h = C_K^f$ for all t)

$$\dot{S} = 2p_G K_A^i(S) - aS \tag{17}$$

$$\dot{C}_K^i = \left(\frac{1-\sigma}{\sigma} \frac{1-\gamma}{\gamma(1-\alpha)} K_A^i(S)^{-1} - \frac{\rho}{\sigma} + C_K^i \right) C_K^i. \tag{18}$$

The eigenvalues in the proximity of the steady state are $EV_1 = C_K^i > 0$ and $EV_2 = -a\frac{1-\gamma-\delta}{1-\gamma} < 0$. EV_2 is negative as we have assumed that the externalities from capital do not outweigh its positive effect on production $(1 - \gamma - \delta > 0)$. As one eigenvalue is negative and our system contains one jump variable and one predetermined variable, the economy is saddle-path stable.

3.2 *Scenario 2: Partial Internalization of the Global Externality*

In contrast to the previous section, we now assume that the regulators are aware of the damages from global pollution. They, however, only internalize the damages from their respective pollution on their own economy while ignoring the effects on the other country. Their intertemporal maximization problem now reads

$$H^{i^2}(C^i, K^i, A_L^i, A_{LG}^i, S, \lambda^i, \mu^i) = H^{i^1}(\cdot) + \mu^i \left(p_G(K_A^i + K_A^j) - aS \right) \tag{19}$$

where $H^{i^1}(\cdot)$ is given in (7) and μ denotes the negative shadow price of stock pollution. Due to the partial internalization of damages from S, the FOCs for A_{LG}^i and K^i are modified while the FOCs for C^i and A_L^i remain unchanged. Together with an additional FOC for the pollution stock, we get after some rearranging

$$\gamma(1-\alpha) Y_K^i K_A^i = 1 + \frac{\mu^i}{\lambda^i} \left(p_G \frac{1}{A_{LG}^i} \right) K_A^i \tag{20}$$

$$g_{\lambda^i} = -(1-\gamma) Y_K^i - \frac{\mu^i}{\lambda^i} \left(p_G \frac{1}{A_{LG}^i} \right) \tag{21}$$

$$g_{\mu^i} = \frac{\lambda^i}{\mu^i} \left(\delta \frac{Y^i}{S} \right) + a \tag{22}$$

plus the additional transversality condition $\lim_{t\to\infty} \mu^i S = 0$.

In comparison to (9), the additional term on the RHS of (20) represents the internalized return to A^i_{LG} from mitigating global pollution. The return is the higher, the higher the negative impact of pollution on welfare relative to the positive welfare effect from capital accumulation, i.e. $\left|\frac{\mu^i}{\lambda^i}\right|$, and the higher the marginal impact of abatement on pollution, i.e. the higher $\left|\frac{\partial P^i_G}{\partial A^i_{LG}}\right| = p_G \frac{K^i_A}{A^i_{LG}}$. Due to the additional return from A^i_{LG}, the optimal ratio of A^i_{LG} to A^i_L rises in comparison to Scenario 1 $\left(\frac{A^i_{LG}}{A^i_L} > \frac{1-\alpha}{\alpha}\right)$. With respect to (10) and (21), the additional term in (21) reflects the internalized costs of capital in terms of global pollution. Finally, (22) gives the dynamics of μ^i, the shadow price of the pollution stock. The growth rate of μ^i is determined by the marginal costs accruing from an additional unit of pollution P^i_G. These costs equal output foregone due to a marginal addition to the pollution stock net of the regeneration rate of S.

From the FOCs we get the modified capital-abatement ratio along the BGP (for the derivation see section "Scenario 2: Derivation of K^i_A and g^i_C Along the BGP" in the Appendix)

$$K^i_A = \left(\gamma(1-\alpha) + \frac{1}{2}\frac{\delta a}{C^i_K + a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma-\delta}} \left(\frac{p_L^\gamma \left(\frac{2p_G}{a}\right)^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{\frac{1}{1-\gamma-\delta}}. \quad (23)$$

Comparison with (16) shows that the equilibrium capital-abatement ratio in (23) is lower due to the internalization of the global externality. The additional term in (23) reflects the reduced incentives to invest in capital due to the negative effect on productivity from global pollution. The stronger this effect is (i.e. the higher δ), the lower the optimal capital-abatement ratio is compared to Scenario 1. From (15) it follows immediately that the pollution stock along the BGP is the lower, the lower the capital-abatement ratio. So, as to be expected, internalization of the global externality induces the global pollution stock to fall.

Regarding local pollution abatement two effects arise due to the internalization of the global externality. On the one hand, we have already noted that compared to Scenario 1, the local versus global abatement ratio falls, as the return to investing in A^i_{LG} rises. This would, *ceteris paribus*, induce lower local pollution abatement, A^i_L, and thus potentially higher local pollution. On the other hand, the internalization of the global externality leads, again *ceteris paribus*, to a lower capital-abatement ratio, K^i_A, as the negative effect of global pollution on productivity is internalized. As we can see from combining (9) and (20), a decline of K^i_A triggers a reduction in the ratio of capital to local abatement and thus lowers local pollution. It can be shown by comparing the equilibrium values of $\frac{K^i}{A^i_L}$ in Scenario 1 and 2 that the abatement increasing effect prevails and equilibrium $\frac{K^i}{A^i_L}$ falls. From (9), (16), (20), (23) and (55) we derive[2]

$$\frac{K^{i(2)}}{A_L^{i(2)}}=\frac{K^{i(1)}}{A_L^{i(1)}}\left(1-\frac{\frac{1}{2}\frac{\delta}{\gamma(1-\alpha)}\frac{a}{C_K^{(2)}+a}}{1+\frac{1}{2}\frac{\delta}{\gamma(1-\alpha)}\frac{\delta a}{C_K^{(2)}+a}}\right)^{-1}\left(1+\frac{1}{2}\frac{\delta}{\gamma(1-\alpha)}\frac{\delta a}{C_K^{(2)}+a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma-\delta}}<1. \tag{24}$$

The first term in brackets on the RHS shows the pollution increasing effect of the crowding-out of local abatement while the second term in brackets reflects the pollution decreasing effect of higher local abatement following higher global abatement. By rearranging terms, it can be shown that $\frac{K^{i(2)}}{A_L^{i(2)}}<\frac{K^{i(1)}}{A_L^{i(1)}}$. As the ratios of capital to local as well as to global abatement fall due to the internalization of the local externality, local pollution decreases compared to a situation in which the regulator only considers the adverse effects of local pollution. So, internalizing global pollution leads to a win-win situation with respect to falling pollution levels locally and globally.

The modified growth rate of consumption is given by (derivation see section "Scenario 2: Derivation of K_A^i and g_C^i Along the BGP" in the Appendix)

$$g_C^i=\frac{1}{\sigma}\left(\left(1-\gamma-\frac{1}{2}\frac{\delta a}{C_K^i+a}\right)Y_K^i-\rho\right). \tag{25}$$

Comparing (25) to (11) shows that internalization of the stock externality has two effects on the growth rate. On the one hand, the equilibrium growth rate is lower due to the reduced marginal return to capital for any given output-capital ratio. The term $\frac{1}{2}\frac{\delta a}{C_K^i+a}$ represents the internalized share of the present value of damages from global pollution which lowers the productivity of capital. On the other hand, changes in the long-run equilibrium output-capital ratio, Y_K^i, also affect the growth rate. From (4) and the FOC for A_L^i in (9), we get Y_K^i as a function of K_A^i only:

$$Y_K^i=\left(p_L^{-\gamma}(\alpha\gamma)^{\alpha\gamma}\left(\frac{2p_G}{a}\right)^{-\delta}\right)^{\frac{1}{1-\alpha\gamma}}(K_A^i)^{-\frac{(1-\alpha)\gamma+\delta}{1-\alpha\gamma}}. \tag{26}$$

Equation (26) is derived using only the production technology and the first-order condition for A_L^i which holds in Scenario 1 as well as in Scenario 2. Consequently, (26) can be used to compare the value of Y_K^i in Scenarios 1 and 2. We see that, as the optimal capital-abatement ratio is lower in Scenario 2 than in Scenario 1, the output-capital ratio is higher. Less output is used in the accumulation of polluting capital and more is spent on abatement activities. That the share of output used for global abatement rises can be shown by multiplying (26) by K_A^i to give Y_A^i, the output-abatement ratio.

[2]Exponents $^{(i)}$, $i=1,2,3,4$ refer to the respective scenarios for comparison purposes.

To see whether the positive or the negative effect of internalization on the growth rate dominates, we rewrite (25) as (for the derivation see section "Scenario 2: Derivation of K_A^i and g_C^i Along the BGP" in the Appendix)

$$g_C^i = \frac{1}{\sigma}\left(\frac{(1-\gamma) - \frac{1}{2}\frac{\delta a}{C_K^i + a}}{\gamma(1-\alpha) + \frac{1}{2}\frac{\delta a}{C_K^i + a}}\frac{1}{K_A^i} - \rho\right). \tag{27}$$

Section "Scenario 2: Derivation of K_A^i and g_C^i Along the BGP" in the Appendix shows that the growth rate under partial internalization of the global externality, (27), is always higher than without internalization, (14). As the drag on growth from global pollution is reduced, the economy grows faster.[3]

Although we have seen that the capital-abatement ratio and the pollution stock are lower than in Sect. 3.1, they are still suboptimal. As neither country internalizes the negative spill-overs of its pollution on the other country, the capital-abatement ratio as well as the pollution stock are still suboptimal as is shown in the next subsection.

3.3 Scenario 3: Full Internalization

It is now assumed that both countries internalize the negative effects of their own pollution not only on the domestic economy but also on the other country. As all market failures are perfectly internalized, the resulting growth path is identical to the welfare-maximal growth path a global social planner would choose. The corresponding Hamiltonian considers the development in both countries and therefore reads

$$H^3(C^i, K^i, A_L^i, A_{LG}^i, S, \lambda^i, \mu) = \sum_i H^{i^1}(\cdot) + \mu\left(p_G(K_A^h + K_A^f) - aS\right). \tag{28}$$

The resulting set of FOCs for each country is identical to the previous section with exception of the FOC for the pollution stock which now reads

$$g_\mu = \frac{\delta}{S}\left(\frac{\mu}{\lambda^h}\frac{1}{A_{LG}^h}\right)^{-1} Y_K^h K_A^h + \frac{\delta}{S}\left(\frac{\mu}{\lambda^f}\frac{1}{A_{LG}^f}\right)^{-1} Y_K^f K_A^f + a. \tag{29}$$

Equation (29) reflects that an increase of the pollution stock induces negative externalities in both countries.

[3]Please note that this unambiguous result is due to the fact that damages have a direct negative effect on production and not, for example, on utility in our model.

Following the same line of reasoning as in the previous scenario, it can be shown that the modified capital-abatement ratio and the growth rate along the BGP are given by

$$K_A^i = \left(\gamma(1-\alpha) + \frac{\delta a}{C_K^i + a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma-\delta}} \left(\frac{p_L^\gamma \left(\frac{2p_G}{a}\right)^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{\frac{1}{1-\gamma-\delta}} \tag{30}$$

and

$$g_C^i = \frac{1}{\sigma}\left(\left(1-\gamma-\frac{\delta a}{C_K^i + a}\right) Y_K^i - \rho\right). \tag{31}$$

Comparison of the term $\frac{\delta a}{C_K^i + a}$ in (30) and (31) to $\frac{1}{2}\frac{\delta a}{C_K^i + a}$ in (23) and (25) shows the effect of integrating foreign damages from domestic pollution. As countries are modelled to be perfectly symmetric, internalization of domestic as well as foreign damages implies exactly a doubling of the effect of internalization compared to a scenario in which only domestic damages are internalized.

Compared to Scenario 1, full internalization again increases growth and reduces local and global pollution.

Compared to Scenario 2, however, matters are more complicated. To disentangle the effects from a rising degree of internalization, let us assume for a moment that C_K^i, the endogenous consumption-capital ratio, is unaffected by full internalization. In this case, a rising degree of internalization would simply lead to the same type of effect as described in Sect. 3.2: Due to the increased internalization of damages, it would be optimal to lower the capital-abatement ratio. Hence, the marginal product of capital would rise and the growth rate would be higher in the new long-run equilibrium. At the same time, the global pollution stock would be lower as the capital-abatement ratio is lower.

However, the ceteris paribus assumption of a constant consumption-capital ratio will not hold. Households will find it optimal to adjust their savings decision due to the change in the marginal product of capital. In case households find it optimal to increase their consumption-capital ratio due to full internalization of the global damages, this induces an effect on the capital-abatement ratio and on growth which goes in the opposite direction than the effect described above.

Whether or not the consumption-capital ratio in Scenario 3 is higher or lower than in Scenario 2, depends crucially on the intertemporal elasticity of substitution, $\frac{1}{\sigma}$. Consider, for example, $\sigma > 1$, i.e. an intertemporal elasticity of substitution below unity.[4] If in this case the marginal productivity of capital rises (for example due to the internalization of global pollution damages), the induced intertemporal

[4]This case represents the majority of empirical estimates of intertemporal substitution elasticities, see e.g. Havránek (2015).

income effect dominates the intertemporal substitution effect. As a result, present consumption rises following the increase of the marginal product of capital and the consumption-capital ratio rises (vice versa for $\sigma < 1$). So, if $\sigma > 1$, households find a higher consumption-capital ratio optimal following an increase in the degree of internalization (see section "Reaction of C_K^i to an Increasing Degree of Internalization of Global Damages" in the Appendix).

But a higher consumption-capital ratio implies a lower present value of marginal global damages: Substituting investment in (polluting) capital by present consumption means that capital accumulation from current production is lower and adds less to the pollution stock. Due to the induced fall of the present value of damages, a higher capital-abatement ratio, K_A^i, is optimal. This effect counteracts the previously described negative effect of internalization on the capital-abatement ratio. The C_K^i-induced effect can, however, never completely offset the negative effect (see section "Reaction of C_K^i to an Increasing Degree of Internalization of Global Damages" in the Appendix), such that overall, global and local pollution always fall due to an increase in the share of internalized damages.

3.4 Scenario 4: Asymmetric Internalization

So far, it was assumed that all countries internalize the global and the local externality symmetrically, yet the current debate on climate policies shows that this is hardly the case. In reality, a number of countries largely ignore global externalities and focus on the internalization of local externalities while others continuously increase their efforts to reduce greenhouse emissions. The widely differing commitments to emission reduction under the Kyoto Protocol reflected this situation quite well. But also under the Paris agreement, the mitigation commitments of many of countries hardly go beyond emission reductions in the status quo, i.e. the level of emission reduction that follows quasi-automatically from technological development and is not driven by domestic climate policy.

In this section we assume that country h internalizes the damages from the local externality and the domestic damages from the global externality while country f solely takes the local externality into account.[5] Under this assumption, the optimization problem of country f is again given by the Hamiltonian of Sect. 3.1 such that its optimal capital-abatement ratio is given by (13)

$$K_A^f = (\gamma(1-\alpha))^{-\frac{1-\alpha\gamma}{1-\gamma}} \left(\frac{p_L^\gamma S^\delta}{(\alpha\gamma)^{\alpha\gamma}} \right)^{\frac{1}{1-\gamma}}.$$

[5] Alternatively, we could also consider all other combinations of internalization scenarios between the two countries, for example, the case in which one country internalizes only domestic externalities from P_G while the other internalizes the global externality perfectly. As, however, the basic implications remain the same, we focus on the above described combination of scenarios.

Assuming that country h takes the repercussions of global pollution at home but not abroad into account, its optimization problem is represented by the Hamiltonian of Sect. 3.2.[6] In this case its capital-abatement ratio is equal to

$$K_A^h = \left(\gamma(1-\alpha) + \frac{\delta a}{C_K^h + a} \frac{K_A^h}{K_A^h + K_A^f}\right)^{-\frac{1-\alpha\gamma}{1-\gamma}} \left(\frac{p_L^\gamma S^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{\frac{1}{1-\gamma}} \tag{32}$$

where $\frac{K_A^h}{K_A^h + K_A^f}$ reflects the share of internalized global damages [that was equal to $\frac{1}{2}$ (resp. 1) in the symmetric Scenario 2 (resp. Scenario 3)].

Equation (32) shows that, compared to Scenario 1, country h finds a lower capital-abatement ratio optimal such that it pollutes less for any given pollution stock. This reduces the equilibrium pollution stock below the equilibrium level of Scenario 1. Due to the decrease in global pollution, country f also lowers its capital-abatement ratio below the mutual non-internalization level. The ratio between the equilibrium global pollution stock under asymmetric internalization, $S^{(4)}$, and without internalization of the global externality, $S^{(1)}$, is given by

$$\left(\frac{S^{(4)}}{S^{(2)}}\right)^{\frac{1-\gamma-\delta}{1-\gamma}} = \frac{1}{2} + \frac{1}{2}\left(1 + \frac{K_A^{h(4)}}{K_A^{h(4)} + K_A^{f(4)}} \frac{\delta}{\gamma(1-\alpha)} \frac{a}{C_K^{h(4)} + a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma}} < 1. \tag{33}$$

The reduction of K_A^i in both countries also leads to lower local pollution in both countries compared to Scenario 1. Again, the effect on local pollution in country f is strengthened by a simultaneous increase in local abatement. In country h, however, higher investment in local abatement is again at least partly crowded out due to the increased focus on global pollution. Yet, it can be shown along the same lines as under Scenario 2 that the net effect on local pollution is still negative.

Comparing the share of the global damages that is internalized in the asymmetric Scenario 4 to the share internalized in the symmetric Scenario 2, we get from (13), (23) and (32)

$$\frac{K_A^{h(4)}}{K_A^{h(4)} + K_A^{f(4)}} = \frac{1}{1 + \left(1 + \frac{K_A^{h(4)}}{K_A^{h(4)} + K_A^{f(4)}} \frac{\delta}{\gamma(1-\alpha)} \frac{a}{C_K^{h(4)} + a}\right)^{\frac{1-\alpha\gamma}{1-\gamma}}} < \frac{1}{2} = \frac{K_A^{h(2)}}{K_A^{h(2)} + K_A^{f(2)}}. \tag{34}$$

[6]Alternatively, we could assume that country h does consider international spill-overs (in which case it would maximize the Hamiltonian of Sect. 3.3). Yet—comparable to the results of Scenarios 2 and 3 presented previously—no additional qualitative effects would arise as only the magnitude of the effects would change.

The ratio between the global pollution stock under Scenario 2 and under Scenario 4 is given by

$$\left(\frac{S^{(4)}}{S^{(2)}}\right)^{\frac{1-\gamma-\delta}{1-\gamma}} = \frac{1}{2}\frac{\left(1+\frac{K_A^{h(4)}}{K_A^{h(4)}+K_A^{f(4)}}\frac{\delta}{\gamma(1-\alpha)}\frac{a}{C_K^{h(4)}+a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma}}+1}{\left(1+\frac{1}{2}\frac{\delta}{\gamma(1-\alpha)}\cdot\frac{a}{C_K^{h(2)}+a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma}}}. \tag{35}$$

Using (34) and considering that a rising degree of internalization can never be overcompensated by a rise of C_K^i (recall Sect. 3.3), we can show:

$$\left(\frac{S^{(4)}}{S^{(2)}}\right)^{\frac{1-\gamma-\delta}{1-\gamma}} > \frac{1}{2}+\frac{1}{2}\left(1+\frac{1}{2}\frac{\delta}{\gamma(1-\alpha)}\frac{a}{C_K^{h(2)}+a}\right)^{\frac{1-\alpha\gamma}{1-\gamma}} > 1 \tag{36}$$

Equivalently, K_A^i and P_L^i are higher and growth is lower under asymmetric internalization than in the symmetric scenarios.

Summing up, asymmetric internalization leaves the foreign country better off than in a scenario with no internalization of the global damages. Triggered by the decrease in the global pollution stock, country f adjusts its capital-abatement ratio and its local abatement such that not only damages from global pollution are decreased but also growth is higher and local pollution is lower. Country h on the other hand is worse off than in a scenario with mutual internalization of global damages where global pollution is lower and higher incentives arise to increase local as well as global abatement.

From the preceding analysis of the different scenarios, the straightforward question arises how the different scenarios could be implemented in a market economy. Although the focus of the paper is on the implications of different internalization scenarios for pollution, abatement and growth, let us take a short look on the design of pollution taxes that could be employed to implement the previously derived solutions. In reality, the policy maker can obviously choose between the implementation of different instruments (e.g. taxes, permit markets, command and control measures). In this paper, however, we stick for simplicity to the analysis of environmental taxation.

4 Market Solution and Environmental Taxation

4.1 Symmetric Scenarios 1–3

Let us first consider optimal policies a policy maker would adopt in the symmetric Scenarios 1–3. As market failures are solely due to environmental externalities, we only have to consider the optimal design of environmental policy.

In the market economy, households maximize their intertemporal utility subject to their intertemporal budget constraint, $\dot{W}^i = r^i W^i - C^i$, where W denotes households' assets. From the FOCs we get the standard Keynes-Ramsey rule

$$g_C^i = \frac{1}{\sigma}\left(r^i - \rho\right). \tag{37}$$

Firms maximize profits which gives rise to FOCs for the two types of abatement and capital. As individual firms do not internalize the externalities arising from pollution, their return to abatement solely consists in the taxes saved due to abatement.

In **Scenario 1**, both countries ignore the global externality such that only the local externality remains to be internalized. Assume that the policy maker levies a tax τ_L on local pollution P_L^i. (In the following we drop country indices for convenience as optimal policies in both countries are identical along the BGP in the symmetric scenarios.) Profit maximization of firms is thus given by

$$\max_{K,A_L,A_{LG}} \Pi^1 = Y(K, P_L(K, A_L, A_{AG}), S) - rK - \tau_L P_L(K, A_L, A_{AG}) - A_L - A_{LG} \tag{38}$$

where the price of output is set to unity and the FOCs for abatement and capital read

$$1 = (1-\alpha)\tau_L \frac{P_L}{A_{LG}} = \alpha \tau_L \frac{P_L}{A_L} \tag{39}$$

$$r = Y_K - \tau_L \frac{P_L}{K}. \tag{40}$$

The optimal policy in this case is given by

$$\tau_L = \gamma \frac{Y}{P_L}, \tag{41}$$

i.e. the optimal tax rate has to equal the marginal externality. Inserting (41) into the above FOCs and the Keynes-Ramsey rule replicates the growth path of Scenario 1. As local pollution is constant along the BGP, it can be seen from (41) that the tax rate has to rise over time in order to mirror the increasing scarcity of pollution in a growing economy.

In **Scenario 2** both regulators additionally internalize the domestic effects of the global externality. Recall that the two externalities can be internalized by two abatement technologies that are not perfect substitutes. In the unregulated state of the world, no market exists for either technology. In order to implement the optimal relative price between the two abatement technologies as well as between abatement and capital, the regulator has to implement two taxes and thereby create a market for each technology.

In addition to the tax on local pollution we now consider a second tax on global pollutant, τ_G, such that firms' profit maximization is given by

$$\max_{K,A_L,A_{LG}} \Pi^2 = \Pi^1 - \tau_G P_G. \tag{42}$$

Due to the additional tax the FOCs for A^i_{LG} and K^i now read

$$1 = (1-\alpha)\tau_L \frac{P_L}{A_{LG}} + \tau_G \frac{P_G}{A_{LG}} \tag{43}$$

$$r = Y_K - \tau_L \frac{P_L}{K} - \tau_G \frac{P_G}{K} \tag{44}$$

while the FOC for A_L remains unchanged. As the local externality was already perfectly internalized in Scenario 1, the optimal tax rule for local pollution is again given by (41). The optimal tax on P_G can then be shown to equal the marginal externality arising from global pollution (see section "Scenario 2: Derivation of Optimal τ_G" in the Appendix)

$$\tau_G = \delta \frac{Y}{S} \frac{1}{C_K + a}. \tag{45}$$

Recall that (41) equalized the tax rate on local pollution, τ_L, to current marginal damages from P_L, $\gamma \frac{Y}{P_L}$. Equivalently, the marginal damage from P_G, i.e. $\delta \frac{Y}{S}$ appears in (45). Yet, the optimal τ_G is also determined by a second term, $\frac{1}{C_K + a}$, that accounts for the present value of current and future damages from today's addition to the pollution stock. The lower this present value of damages, the lower the optimal tax rate will be that internalizes these damages. The tax rate depends therefore negatively on the regeneration rate a, as faster regeneration implies that pollution is absorbed faster. Consequently, the present value of current emissions will be lower for a higher a. Also, a higher consumption-capital ratio, as described in Sect. 3.3, implies a lower present value of damages from global pollution.

Finally, let us consider **Scenario 3**. As only the scope of internalization changes compared to Scenario 2, the same instruments can be employed. Regarding the local externality, (41) still represents the optimal tax rule while (45) has to be adjusted in order to capture the international spill-overs from global pollution. Recalling the results for Scenario 3, it follows straightforwardly that the optimal τ_G in Scenario 3 is given by

$$\tau_G = 2\delta \frac{Y}{S} \frac{1}{C_K + a}. \tag{46}$$

4.2 Asymmetric Scenario 4

Given that the two countries do not follow the same internalization strategy in Scenario 4, the question arises which policy mix is optimal in this case. Given our basic assumption that both countries take the pollution that is generated in the other country as exogenous, it follows straightforwardly that the optimal policy rules are identical to those derived in the previous subsection. More specifically, both countries set the local pollution tax according to (41) and country h taxes P_G according to (45) if it only takes the domestic damages from global pollution into account or (46) if it also considers the damages in country f. Yet, although the policy rules are the same as in Sect. 4.1, the absolute levels and growth rates of taxation are different. As the stock of pollution in Scenario 4 is neither identical to Scenario 1 nor to Scenario 2 (or 3), the growth rate and the level of economic activities also differ from Scenarios 1 and 2 (or 3) which entails different levels and growth rates of the optimal tax rates.

To show that it cannot be optimal for country h to choose different policy rules, consider the following: Let us assume that country h is aware that country f only internalizes its local damages. In this case, country h is also aware that, despite its internalization efforts, overall global pollution is still above the level that would be optimal if both countries internalized global damages. This case probably constitutes a quite realistic representation of the current political situation: Some countries are actively pursuing climate policy while others concentrate mostly on the reduction of, for example, local air pollution. As a consequence, global mitigation efforts are not sufficient to curb climate change significantly which the countries active in climate policy are well aware of.

In order to reach the optimal global pollution stock of Scenario 2 (or 3), country h could tax global pollution at a higher rate, thus inducing further abatement which would yield an accompanying reduction of local pollution as ancillary benefit. This policy can, however, not be optimal. Under the policy rules we derived in the previous sections, country h sets the tax rate, τ_G, equal to the marginal damage from pollution. For a τ_G above this level, the marginal damages would be smaller than the tax rate which would lower domestic welfare. Of course, the country could—as compensation for the higher τ_G—lower the tax on local pollution. By raising τ_G beyond the level specified by (46) and simultaneously lowering τ_L below the level of (41), regulation could lead to the optimal local and global pollution levels. The costs of attaining these optimal levels would, however, be suboptimally high as the price ratio between the two types of pollution would be distorted.

5 Conclusions

This paper has analyzed the impact of different scopes of internalization of global and local pollution on the long-run development of economies as well as on the development of pollution. The analysis considered two countries that share the same global environment and suffer from individual local pollution. It was assumed that the countries each have two abatement technologies at their disposal. One technology only mitigates local pollution while the other one also reduces global pollution.

The motivation for the paper mainly stemmed from the observation that the focus of environmental policies differs quite substantially between countries. While some place the emphasis on local pollution abatement, e.g. from air or water pollution, others focus more on long-run global pollution problems, e.g. climate change. The reasons for these different perspectives can be manifold and are not at the core of this paper. Instead, the paper focused on the consequences of the different internalization scenarios for pollution, growth and the incentives to invest in capital accumulation and/or abatement.

Three different internalization scenarios were considered in which both countries have the same scope of internalization (for example, both internalize local and global pollution damages). In a final scenario, we looked at a situation in which only one of the countries aims at global pollution internalization. Of course, the number of potential scenarios is much higher. Ranging from complete disregard of pollution internalization in either country to perfect internalization of local and global externalities, sixteen different scenarios could have been considered. The additional insights from each scenario comparison would, however, have been quite low. So, we focused on the four scenarios described above that gave insights into fundamental regularities resulting from different scopes of internalization:

- The internalization of local pollution damages not only lowers local but also global pollution if at least one of the abatement technologies employed also reduces global pollution.
- If both countries additionally internalize damages from global pollution, the global pollution stock falls further. This reduction is accompanied by a decline of local pollution as also the marginal benefits from local pollution abatement increase, thus creating a win-win situation.
- The extent to which global pollution falls and growth rises due to an increasing degree of internalization, depends crucially on the savings reaction of households to internalization. If households increase consumption today and thus lower their capital accumulation, the *ceteris paribus* positive effect of internalization on global pollution is partially offset. The increase in the consumption-capital ratio induces a decline in the present value of global pollution and thus reduces incentives to invest in pollution reduction. Vice versa, if households substitute savings for present consumption, the effects of internalization on local and global pollution reduction are strengthened.
- The effects of internalization (a) on investment in abatement compared to investment in capital and (b) on economic growth mirror the just described effects on pollution. The internalization of the global externality increases incentives to substitute capital by abatement and the reduction of the productivity decreasing effects of pollution boosts growth.

- Asymmetric internalization induces abatement and pollution levels that are between the levels for the respective symmetric internalization scenarios. The level of abatement increases not only in the country with the wider scope of internalization but also in the other country. Triggered by the improved environmental quality and its productivity enhancing effects, it pays off for both countries to spend more on local as well as global abatement. This transmission channel of global environmental policy is often ignored in the literature.
- To implement the different internalization scenarios, taxes on pollution can be employed. Two taxes are required to internalize the two externalities perfectly. As to be expected, the tax on local pollution reflects the damages from a marginal increase of the flow of local pollution while the global pollution tax reflects the present value of today's and future damages from a marginal addition to the pollution stock. The general tax design is independent of the extent of internalization of damages and of the internalization strategy.

The focus of this paper was to provide a first take on fundamental mechanisms induced by different scopes of internalization in the presence of pollution of differing regional reach. For this first take, we have made a number of simplifying assumptions (perfect information, symmetric countries and non-strategic behavior). Going beyond our analysis by loosening either of these assumptions offers interesting scope for future research and further interesting results.

Appendix

Scenario 2: Derivation of K_A^i and g_C^i Along the BGP

From the FOCs for A_{LG}^i and K^i, (20) and (21), we get

$$g_{\lambda^i} = -(1-\alpha\gamma)Y_K^i + {K_A^i}^{-1} \tag{47}$$

while combining (20) and (22) gives

$$g_{\mu^i} = \delta \frac{1}{S} P_G^i Y_K^i \left(\gamma(1-\alpha)Y_K^i - {K_A^i}^{-1}\right)^{-1} + a. \tag{48}$$

From differentiating (20) with respect to time, we get a second expression for the dynamics of g_{μ^i}

$$g_{\mu^i} = g_{\lambda^i} + g_{A_{LG}^i} - g_K^i + \frac{\gamma(1-\alpha)g^i_{Y}\frac{Y^i}{A_{LG}^i} - g_{A_{LG}^i}}{\gamma(1-\alpha)\frac{Y^i}{A_{LG}^i} - 1}. \tag{49}$$

Along the BGP $g_{C^i} = g_{Y^i} = g_{K^i} = g_{A_L{}^i} = g_{A^i_{LG}}$ again has to hold, such that we get from (49) that along the BGP $g_{\mu^i} = g_{\lambda^i} + g_{K^i}$. Using also

$$g_{K^i} = (1 - \alpha\gamma)Y^i_K - C^i_K - K^{i^{-1}}_A \tag{50}$$

from (5) where we employed (9), we get

$$g_{\mu^i} = -C^i_K \tag{51}$$

and from equating (48) and (51)

$$-(C^i_K + a)\frac{S}{\delta P^i_G} = \left(\gamma(1-\alpha) - (K^i_A Y^i_K)^{-1}\right)^{-1}. \tag{52}$$

Using (26) gives

$$K^i_A = \left(\gamma(1-\alpha) + \frac{\delta a}{C^i_K + a}\frac{K^i_A}{K^i_A + K^j_A}\right)^{-\frac{1-\alpha\gamma}{1-\gamma}} \left(\frac{p^\gamma_L S^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{\frac{1}{1-\gamma}}. \tag{53}$$

Finally, by employing (15) we can now derive (23), the capital-abatement ratio along the BGP.

To derive the optimal output-capital ratio rearrange (20) to get

$$Y^i_K K^i_A = (\gamma(1-\alpha))^{-1}\left(\frac{\mu^i}{\lambda^i}\left(p_G\frac{K^i_A}{A^i_{LG}}\right) + 1\right). \tag{54}$$

Inserting this expression into (52) gives

$$\frac{\mu^i}{\lambda^i}\left(p_G\frac{K^i_A}{A^i_{LG}}\right) = -\frac{\frac{1}{2}\frac{\delta a}{C^i_K+a}}{\frac{1}{2}\frac{\delta a}{C^i_K+a} + \gamma(1-\alpha)} \tag{55}$$

such that Y^i_K can be rewritten as

$$Y^i_K = \frac{1}{K^i_A}\left[\frac{1}{\frac{1}{2}\frac{\delta a}{C^i_K+a} + \gamma(1-\alpha)}\right]. \tag{56}$$

The BGP growth rate in Scenario 2 can be derived from (8), (20) and (21) to equal

$$g^i_C = \frac{1}{\sigma}\left((1-\alpha\gamma)Y^i_K - (K^i_A)^{-1} - \rho\right). \tag{57}$$

Using (56) we get (25) and (27) respectively.

To show that the growth rate in Scenario 2, (27), is higher than in Scenario 1, (14), consider the following: Inserting K_A^i from (16) and (23) into (14) and (27) respectively shows that $g_C^{i(27)} > g_C^{i(14)}$ holds for

$$A > 1 \text{ with } A = \left(1 - \frac{1}{2}\frac{1}{1-\gamma}\frac{\delta a}{C_K^i + a}\right)^{1-\gamma-\delta}\left(1 + \frac{1}{2}\frac{1}{\gamma(1-\alpha)}\frac{\delta a}{C_K^i + a}\right)^{\gamma(1-\alpha)+\delta}$$

Remembering that $0 < \delta < (1-\gamma)$, we get

$$\lim_{\delta \to 0} A = 1$$

$$\lim_{\delta \to (1-\gamma)} A = \left(1 + \frac{1}{2}\frac{1-\gamma}{\gamma(1-\alpha)}\frac{\delta a}{C_K^i + a}\right)^{\gamma(1-\alpha)} > 1$$

with A rising monotonously in δ for $\delta \in [0, 1-\gamma]$:

$$\frac{\partial A}{\partial \delta} = \frac{A \cdot \log\left[\left(\frac{1+\frac{1}{2}\frac{1}{\gamma(1-\alpha)}\frac{\delta a}{C_K^i+a}}{1-\frac{1}{2}\frac{1}{1-\gamma}\frac{\delta a}{C_K^i+a}}\right)^{\left(\gamma(1-\alpha)-\frac{1}{2}\frac{\delta a}{C_K^i+a}\right)\left(1-\gamma+\frac{1}{2}\frac{\delta a}{C_K^i+a}\right)}\right]}{\left(1-\gamma-\frac{1}{2}\frac{1}{1-\gamma}\frac{\delta a}{C_K^i+a}\right)\left(\gamma(1-\alpha)+\frac{1}{2}\frac{\delta a}{C_K^i+a}\right)} > 0.$$

As $A > 1$ holds for $0 < \delta < 1-\gamma$, internalization of the global externality raises the growth rate (i.e. $g_C^{i(27)} > g_C^{i(14)}$).

Reaction of C_K^i to an Increasing Degree of Internalization of Global Damages

To derive the effect an increasing degree of internalization of the global externality has in the symmetric scenarios, consider the following slightly more general version of (30)

$$K_A^i = \left(\gamma(1-\alpha) + \Delta\frac{\delta a}{C_K^i + a}\right)^{-\frac{1-\alpha\gamma}{1-\gamma-\delta}}\left(\frac{p_L^\gamma\left(\frac{2p_G}{a}\right)^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{\frac{1}{1-\gamma-\delta}} \tag{58}$$

where $0 < \Delta < 1$ represents the degree of symmetric internalization of the global damages (recall that for Scenario 2, $\Delta = \frac{1}{2}$, while for Scenario 3, $\Delta = 1$, holds).

Inserting (58) into (26) gives

$$Y_K^i = \left(\gamma(1-\alpha) + \Delta\frac{\delta a}{C_K^i + a}\right)^{\frac{(1-\alpha)\gamma+\delta}{1-\gamma-\delta}} \left(\frac{p_L^\gamma\left(\frac{2p_G}{a}\right)^\delta}{(\alpha\gamma)^{\alpha\gamma}}\right)^{-\frac{1}{1-\gamma-\delta}}. \tag{59}$$

From equating

$$g_C^i = \frac{1}{\sigma}\left(\left(1-\gamma-\Delta\frac{\delta a}{C_K^i + a}\right)Y_K^i - \rho\right) \tag{60}$$

which represents the BGP growth rate (31) for a degree of internalization of Δ, to (50) under consideration of $g_{C^i} = g_{K^i}$, we get C_K^i as a function of parameters only after inserting (56) and (59), where $\frac{1}{2}$ in (56) was substituted by $\frac{1}{\Delta}$:

$$C_K^i = \frac{\rho}{\sigma} + \frac{\sigma-1}{\sigma}\left(1-\gamma-\Delta\frac{\delta a}{C_K^i + a}\right)\left(\gamma(1-\alpha) + \Delta\frac{\delta a}{C_K^i + a}\right)^{\frac{\gamma(1-\alpha)+\delta}{1-\gamma-\delta}}\Omega \tag{61}$$

with $\Omega = \left(p_L^\gamma\left(\frac{2p_G}{a}\right)^\delta(\alpha\gamma)^{-\alpha\gamma}\right)^{-\frac{1}{1-\gamma-\delta}}$.

The reaction of the consumption-capital ratio to a marginal increase in the degree of internalization is then given by

$$\frac{\partial LHS}{\partial C_K^i}dC_K^i = \frac{\partial RHS}{\partial C_K^i}dC_K^i + \frac{\partial RHS}{\partial \Delta}d\Delta. \tag{62}$$

As $\frac{\partial LHS}{\partial C_K^i} = 1$, we get

$$\frac{dC_K^i}{d\Delta} = \frac{\partial RHS}{\partial \Delta}\left(1 - \frac{\partial RHS}{\partial C_K^i}\right)^{-1} \tag{63}$$

where

$$\frac{\partial RHS}{\partial C_K^i} = -\frac{\sigma-1}{\sigma}\left[1-\Delta\frac{a}{C_K^i + a}\right]\Theta \tag{64}$$

$$\frac{\partial RHS}{\partial \Delta} = -\frac{C_K^i + a}{\Delta}\frac{\partial RHS}{\partial C_K^i} \tag{65}$$

with $\Theta = \frac{(1-\alpha\gamma)\delta}{1-\gamma-\delta}\Omega\left(\gamma(1-\alpha)+\Delta\frac{\delta a}{C_K+a}\right)^{\frac{\gamma(1-\alpha)+\delta}{1-\gamma-\delta}-1}\left(\Delta\frac{\delta a}{(C_K+a)^2}\right) > 0$. Collecting terms we get

$$\frac{dC_K^i}{d\Delta} = \frac{C_K^i+a}{\Delta}\frac{\frac{\sigma-1}{\sigma}\left[1-\Delta\frac{a}{C_K^i+a}\right]\Theta}{1+\frac{\sigma-1}{\sigma}\left[1-\Delta\frac{a}{C_K^i+a}\right]\Theta}. \tag{66}$$

The sign of (66) clearly depends on $\sigma \gtrless 1$: For $\sigma > 1$, we get an unambiguous increase in C_K^i for a marginal increase in Δ. For $\sigma < 1$, C_K^i can increase or decrease depending on the parametrization of the model.

To see whether $\Delta\frac{\delta a}{C_K^i(\Delta)+a}$ could decrease following an increase in Δ, consider that in this case

$$\frac{d\frac{\Delta\delta a}{C_K^i+a}}{d\Delta} < 0 \qquad \Leftrightarrow \qquad \frac{dC_K^i}{d\Delta} > \frac{(C_K^i+a)}{\Delta} \tag{67}$$

would have to hold. Comparison with (66) shows that this condition would only be met for $1+\frac{\sigma-1}{\sigma}\left[1-\Delta\frac{a}{C_K^i+a}\right]\Omega < 0$. In this case, however, $\frac{\partial RHS}{\partial C_K^i} > 1$ in (61) such that (61) would have no interior solution for C_K^i in the long-run equilibrium.

Having determined that $\Delta\frac{\delta a}{C_K^i(\Delta)+a}$ is unambiguously higher for a higher degree of internalization, it follows directly from (58) that the capital-abatement ratio is lower. From (59) and (60) we get for the reaction of the growth rate to an increase in the degree of internalization:

$$\frac{dg_C^i}{d\Delta} = \frac{\Theta}{\sigma(1-\alpha\gamma)}\left(\Delta\frac{\delta a}{(C_K^i+a)^2}\right)^{-1}\left[1-\Delta\frac{a}{C_K^i+a}\right]\frac{d\left(\Delta\frac{\delta a}{C_K^i(\Delta)+a}\right)}{d\Delta} > 0. \tag{68}$$

For local pollution, $P_L^i = p_L\left(\frac{K^i}{A_L^i}\right)^{\alpha}(K_A^i)^{1-\alpha}$ we get from the FOC for A_L^i, (9),

$$\frac{d\frac{K^i}{A_L^i}}{d\Delta} = -\frac{(Y_K^i)^{-2}}{\alpha\gamma}\frac{\partial Y_K^i}{\partial K_A^i}\frac{dK_A^i}{d\left(\Delta\frac{\delta a}{C_K^i(\Delta)+a}\right)}\left[1-\Delta\frac{a}{C_K^i+a}\right]\frac{d\left(\Delta\frac{\delta a}{C_K^i(\Delta)+a}\right)}{d\Delta} < 0. \tag{69}$$

As K_A^i and $\frac{K^i}{A_L^i}$ both decrease, local pollution falls unambiguously.

Scenario 2: Derivation of Optimal τ_G

To determine the optimal τ_G, first insert τ_L from (41) into (43) which gives

$$\tau_G \frac{P_G}{A_{LG}} = 1 - (1-\alpha)\gamma Y_K K_A. \tag{70}$$

From (20) we know that $1 - (1-\alpha)\gamma Y_K K_A = -\frac{\mu}{\lambda}\frac{P_G}{A_{LG}}$ has to hold in the optimum. Equating the two expressions shows that the tax rate has to equal the negative ratio of the shadow prices of stock pollution and capital,

$$\tau_G = -\frac{\mu}{\lambda}. \tag{71}$$

Equating the two expressions for g_μ from (22) and (50) gives

$$-C_K = \frac{\delta}{S}\left(\frac{\mu}{\lambda}\frac{1}{A_{LG}}\right)^{-1} Y_K K_A + a \tag{72}$$

which reads after some rearranging

$$-\frac{\mu}{\lambda} = \delta \frac{Y}{S}\frac{1}{C_K + a}. \tag{73}$$

Combining (71) and (73) finally gives the optimal tax rate τ_G in (45).

References

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy, 97*, 1447–1458.

Auld, D. A. L., & Eden, L. (1990). Public characteristics of non-public goods. *Public Finance, 3*, 378–391.

Bahn, O., & Leach, A. (2008). The secondary benefits of climate change mitigation: An overlapping generations approach. *Computational Management Science, 5*, 233–257.

Cornes, R. (2016). Aggregative environmental games. *Environmental and Resource Economics, 63*, 339–365.

Cornes, R. C., & Sandler, T. (1984). Easy riders, joint production, and public goods. *Economic Journal, 94*, 580–598.

Cornes, R. C., & Sandler, T. (1994). The comparative static properties of the impure public good model. *Journal of Public Economics, 54*, 403–421.

Finus, M., & Rübbelke, D. T. G. (2013). Public good provision and ancillary benefits: The case of climate agreements. *Environmental and Resource Economics, 56*, 211–226.

Gielen, D., & Chen, C. (2001). The CO_2 emission reduction benefits of Chinese energy policies and environmental policies: A case study for Shanghai, period 1995–2020. *Ecological Economics, 39*, 257–270.

Gorman, W. M. (1980). A possible procedure for analysing quality differentials in the egg market. *The Review of Economic Studies, 47*, 843–856.

Grimaud, A., & Rougé, L. (2003). Non-renewable resources and growth with vertical innovations: Optimum, equilibrium and economic policies. *Journal of Environmental Economics and Management, 45*, 433–453.

Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choice and selective reporting. *Journal of the European Economic Association, 13*, 1180–1204.

IPCC. (1996). *Climate change 1995 - Economic and social dimensions of climate change*. Cambridge: Cambridge University Press.

IPCC. (2001). *Climate change 2001 - Mitigation*. Cambridge: Cambridge University Press.

Kotchen, M. J. (2005). Impure public goods and the comparative statics of environmentally friendly consumption. *Journal of Environmental Economics and Management, 49*, 281–300.

Kotchen, M. J. (2006). Green markets and private provision of public goods. *Journal of Political Economy, 114*, 816–834.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*, 132–157.

Lancaster, K. J. (1971). *Consumer demand: A new approach*. New York: Columbia University Press.

Li, J. C. (2006). A multi-period analysis of a carbon tax including local health feedback: An application to Thailand. *Environment and Development Economics, 11*, 317–342.

Morgenstern, R. D. (1991). Towards a comprehensive approach to global climate change mitigation. *The American Economic Review, 81*, 140–145.

Pearce, D. (1992). *Secondary Benefits of Greenhouse Gas Control*. CSERGE Working Paper 92-12, London.

Pearce, D. (2000). Policy framework for the ancillary benefits of climate change policies. In *Ancillary benefits and costs of greenhouse gas mitigation* (pp. 517–560). Paris: OECD.

Pittel, K., & Bretschger, L. (2009). *The Implications of Heterogeneous Resource Intensities on Technical Change and Growth*. CER-ETH Working-Paper 09/120, ETH Zürich.

Pittel, K., & Rübbelke, D. T. G. (2006). What directs a terrorist? *Defence and Peace Economics, 17*, 311–328.

Pittel, K., & Rübbelke, D. T. G. (2008). Climate policy and ancillary benefits. A survey and integration into the modelling of international negotiations on climate change. *Ecological Economics, 68*, 210–220.

Plambeck, E. L., Hope, C., & Anderson, J. (1997). The Page 95 Model: Integrating the science and economics of global warming. *Energy Economics, 19*, 77–101.

Posnett, J., & Sandler, T. (1986). Joint supply and the finance of charitable activity. *Public Finance Quarterly, 14*, 209–222.

Pugliese, T., & Wagner, J. (2011). Competing impure public goods and the sustainability of the theater arts. *Economics Bulletin, 31*, 1295–1303.

Rebelo, S. T. (1991). Long-run policy analysis and long-run growth. *Journal of Political Economy, 99*, 500–521.

Rive, N., & Rübbelke, D. T. G. (2010). International environmental policy and poverty alleviation. *Review of World Economics, 146*, 515–543.

Rübbelke, D. T. G. (2002). *International climate policy to combat global warming - An analysis of the ancillary benefits of reducing carbon emissions. New horizons in environmental economics*. Cheltenham, UK; Northampton, MA: Edward Elgar.

Rübbelke, D. T. G. (2003). An analysis of differing abatement incentives. *Resource and Energy Economics, 25*, 269–295.

Rübbelke, D. T. G. (2005). Differing motivations for terrorism. *Defence and Peace Economics, 16*, 19–27.

Sandler, T. (1996). A game-theoretic analysis of carbon emissions. In R. D. Congleton (Ed.), *The political economy of environmental protection - Analysis and evidence* (pp. 251–272). Ann Arbor : The University of Michigan Press.

Sandler, T., & Hartley, K. (2001). Economics of alliances: The lessons for collective action. *Journal of Economic Literature, 39*, 869–896.
Sandler, T., & Murdoch, J. C. (1990). Nash-Cournot or Lindahl behavior? An empirical test for the NATO allies. *Quarterly Journal of Economics, 105*, 875–894.
Schou, P. (2000). Polluting non-renewable resources and growth. *Environmental and Resource Economics, 16*, 211–227.
Schou, P. (2002). When environmental policy is superfluous: Growth and polluting resources. *Scandinavian Journal of Economics, 104*, 605–620.
Smulders, S., & Gradus, R. (1996). Pollution abatement and long-term growth. *European Journal of Political Economy, 12*, 505–532.
Vicary, S. (1997). Joint production and the private provision of public goods. *Journal of Public Economics, 63*, 429–445.
Withagen, C. (1995). Pollution, abatement and balanced growth. *Environmental and Resource Economics, 5*, 1–8.

Showdown in Schönau: A Contest Case Study

Bouwe R. Dijkstra and Patrick R. Graichen

1 Introduction

Initiated by Tullock (1967), the model of the rent seeking contest is currently a quite popular subject of theoretical analysis. Performing empirical research is difficult due to the twilight in which many attempts to influence political decisions take place. As a result, the sparse empirical research into contests has mainly been limited to indirect effects.

In the absence of a sound empirical basis, theoretical analysis operates mainly with contest success functions like the Tullock (1980) function, because these functions are conventionally used and their properties are convenient and well-known. Furthermore, it is common for theoretical extensions to be made without reference to empirical research showing the relevance of the extension being undertaken.

The present paper addresses this unfortunate situation by presenting a case study of a political contest. Of course, a case study like this cannot be used to test the contest model or to estimate all its parameters. But at least we can try to describe the case in terms of contest theory: payoff, stake, effort, success probability and

The opinions expressed in this paper are entirely the autors' and do not necessarily reflect the views of the German government. The authors would like to thank the interview partners in Schönau for their cooperation and Andries Nentjes, Till Requate and Bert Schoonbeek for helpful comments.

B.R. Dijkstra (✉)
University of Nottingham, Room C3A Sir Clive Granger Building, University Park, Nottingham, NG7 2RD UK
e-mail: Bouwe.Dijkstra@nottingham.ac.uk

P.R. Graichen
Agora Energiewende, Anna-Louisa-Karsch-Straße 2, 10178 Berlin, Germany
e-mail: patrick.graichen@agora-energiewende.de

W. Buchholz, D. Rübbelke (eds.), *The Theory of Externalities and Public Goods*,
DOI 10.1007/978-3-319-49442-5_15

contest success function. We may gain some insight into the nature, the size and the effectiveness of the lobbies' activities, and their decisions about how much and what to do. Finally, we may identify some elements of the actual contest that have not been theoretically modeled yet.

The contest we study is a conflict between environmentalists and an electricity supplier in the small German town of Schönau. Starting in 1986, the environmentalists of Schönau organized a campaign for a more environmentally friendly form of energy production and finally succeeded (after two local referenda) in July 1997 when their own energy supply firm replaced the original supplier. In this chapter, we shall mainly restrict ourselves to an analysis of the second referendum.[1]

Of all political decisions, the referendum is probably the one most accessible to empirical contest research. Attempts undertaken by lobbies to influence a decision made by politicians or bureaucrats are more elusive. Because the actors tend to be secretive about the influence attempts, it is difficult to find out when the major decisions were made and how (hard) interest groups tried to influence these decisions. Furthermore, unlike an election, a referendum is about a single issue.

The rest of the paper is organized as follows. The Schönau story is presented in Sect. 2. In Sect. 3 we introduce our method: the theoretical background of contest theory and our sources. We also review previous empirical research into rent seeking contests. Section 4 discusses the qualitative aspects of the contests: who the active lobbies were, what they did, which arguments they used and what their strategies were. Section 5 addresses the quantitative aspects: stakes, effort, lobbying effectiveness and success probability. In Sect. 6, we analyze the outcome of the referendum. We examine which voter groups were more inclined to vote one way or the other and which factors influenced, or even determined, the outcome. Section 7 presents some implications of our case study for contest models. Section 8 concludes the paper.

2 The Story

After the Chernobyl nuclear accident in April 1986, the anti-atomic movement received widespread support everywhere in Europe. Even in Schönau, a little town of 2500 inhabitants located in the Black Forest (Southern Germany), living mainly on tourism and with a strong Catholic-conservative background, a group called "Parents for a nuclear-free future" was founded.

Starting out as a self-help group, the members soon decided they should try and make an active difference themselves by contributing to a reduction in energy use and to more environmentally-friendly methods of electricity generation. They approached the regional energy monopolist KWR (Kraftübertragungswerke Rheinfelden) asking for a linear tariff structure, the discontinuation of electricity

[1] Graichen (2003) analyzes the whole conflict.

from nuclear power plants (40 % of the electricity came from nuclear power plants) and higher feed-in tariffs for combined heat and power (CHP). KWR refused to discuss these terms. Thus, the action group had to operate on its own. Over the next few years they initiated several courses of action: hearings with experts on the use of renewable energy and CHP in Schönau, an annual energy saving campaign, private recycling initiatives and a benefit concert for the children of Chernobyl, who were also invited to spend their holidays in the Black Forest region.

It was in 1990 that the Schönau case started to develop in an unusual way. Schönau's monopoly concession contract with KWR was not due to expire until 1994, but KWR offered the town a contract renewal from early 1991 for the duration of 20 years, together with a 100,000 DM[2] increase in concession fees.

The environmentalists realized that if they did not act now, they would be saddled with the uncongenial KWR for another 20 years. So they raised 100,000 DM themselves, offering to pay the amount to the town if the contract with KWR was not renewed prematurely. By 1994 they would have set up an energy firm themselves and then the town could choose between KWR and them.

In July 1991 the council voted 7 to 6 for the acceptance of KWR's offer. The 5 CDU (conservative) members, the CDU mayor and one SPD (social democrat) member were in favour, the 4 FWV (independents) members and two SPD members were against.

However, since the constitution of the state of Baden-Württemberg allows for local referenda,[3] the environmentalists rallied for a referendum to suspend the council decision. The referendum was held in October 1991 after three months of intensive campaigning by both sides. The political battle was quite heated and led to a relatively high turnout of 74.3 % with 55.7 % (729 votes) in favour of termination of the contract with KWR.

In the years 1992–1994, the environmentalists launched several operations to prepare for the "final battle". They founded their own electricity firm EWS (Elektrizitätswerke Schönau). Experts devised an energy concept for the town on the basis of regional conditions and green preferences. Money had to be raised, since they would have to buy the electricity grids from KWR.

Two local elections reversed the majority in the town council. The 1993 elections for mayor were won by CDU-backed candidate Seger, who promised to remain neutral on the electricity grant issue. His opponent had spoken out in favour of EWS. In the 1994 town council elections, the FWV (independents) won a seat off the CDU (conservatives).

---

[2] 1 DM (Deutsche Mark) = 0.51 Euro.

[3] The rule is as follows: a referendum to withdraw a town council decision must be held if within 4 weeks after the decision in question 15 % of the voting population signs a referendum claim. The referendum itself results in the rejection of the town council decision if there is a majority in favour of rejection and this majority comprises at least 30 % of the electorate. The town council is then obliged to act accordingly for three years.

Early 1995, both the regional firm KWR and the environmentalists' firm EWS presented a contract offer to the town. The town council finally decided on 20 November 1995 with 6 to 5 votes[4] to grant the electricity concession to the "green" EWS firm. A "Citizens' Pro-Referendum Initiative"[5] was immediately formed, initiated by the local conservative CDU establishment. This group collected enough signatures for a second referendum, to be held in March 1996.

In the winter of 1996, the little town with its voting population of 1800 experienced a political battle of previously unknown vehemence. In this paper, we shall present an extensive account of this battle of the environmentalists against the electricity firm and its local allies. On March 10, 1996 the referendum had the highest electoral turnout ever in the history of Schönau with 84.3 %; 52.4 % (782) voted in favour of the environmentalists, confirming the town council vote from November 1995.

After strenuous negotiations about the price, the environmentalists' firm EWS finally took over the electricity net on 1 July 1997. On that day, of course, little changed outwardly in Schönau. Initially, EWS simply distributed the electricity produced by KWR to the 1700 households of Schönau. But already in the first year, EWS changed a lot: A new energy tariff structure was presented, increasing the incentives to save energy. The energy bill is delivered monthly, so as to provide direct feedback on households' energy consumption. Several small-scale Combined Heat and Power installations and photovoltaic systems were installed in private households and on municipal and church buildings. In late 1998, EWS made use of the liberalization of the European electricity market to replace the nuclear power they bought previously from KWR by hydro-electricity from Austria. At the same time, EWS became a national broker in green electricity and is today supplying over 150,000 households all over Germany with green power.[6]

3 Method

3.1 The Application of Contest Theory

3.1.1 Limitation to the Second Referendum

We analyze the conflict between the environmentalists and the energy firm KWR as a contest. The antagonists are pictured as agents expending resources to try and win the concession. The concession is granted as a result of a multi-stage

[4] For EWS: 4 FWV representatives and two SPD representatives. For KWR: 4 CDU representatives and one SPD representative. The mayor abstained. One FWV representative was not allowed to vote because of his sizeable financial interest in EWS.

[5] We shall refer to this group later as Citizens' Initiative or CI.

[6] See www.ews-schoenau.de.

political decision.[7] First the town council decides who will get the concession. The lobby that has lost in the town council can then try to collect enough support for a referendum. At the referendum, the town council decision is repealed if the majority of votes is for withdrawal and either the turnout or the (absolute) number of votes for withdrawal is high enough.

In the case of Schönau, however, the first two stages of the game did not leave much scope for strategic considerations. Collecting enough signatures for a referendum was not a problem. On each occasion, more than twice the required number were collected. In 1991, the voting behaviour of the town council members was determined beyond influence attempts well in advance. In 1996, the council voted according to the 1991 patterns, with the mayor abstaining as announced. Thus, we do not lose much if we limit the analysis to the referendum.

As we have seen in the previous section, there were two referenda in Schönau about the electricity concession. In this paper, we shall limit ourselves to the second referendum. The second referendum is more interesting because it was final, and therefore the lobbies' stakes were higher. Furthermore, the consequences of rejection or vindication of the town council vote were relatively clear. As a practical matter, there was also more information available about the second referendum.

3.1.2 The Contest for the Referendum

Three lobbying groups were active in the 1996 referendum campaign: the environmentalists on the one side, and the electricity firm KWR plus the Citizens' Initiative (CI) on the other side. In this subsection we discuss the theoretical background of the lobbies' behaviour in terms of contest theory.

In a contest, the players expend resources in order to increase the probability that a certain process will have a favourable outcome.[8] Instances are of R&D races, sports matches or political decisions.

Since a referendum as such has not been modeled, we take the general formulation of the contest model as our starting point.[9,10] The success probability p for the environmentalists can be written as a function of the efforts by all agents:

---

[7]In this paper, we take the content of the concession offers as given. This subject is discussed by Graichen et al. (2001), who show that when challenged by environmentalists, the monopolist's offer will be more environmentally friendly than otherwise. Liston-Heyes (2001) derives similar results.

[8]See Long (2013) for an overview and Congleton et al. (2008) for an anthology.

[9]For an analysis of the general model, see e.g., Hillman (1989), Baik (1994) and Nti (1999). A logit-function contest is an example of an aggregative game (Cornes and Hartley 2007): A player's payoff only depends on his own input and the sum of everyone else's input. Cornes and Hartley (2005, 2012) apply the apparatus of aggregative games to study equilibrium existence, uniqueness and rent dissipation for contests with risk-neutral and risk-averse players, respectively.

[10]We assume that the turnout will be so high that the referendum is valid. Herrera and Mattozzi (2010) show how a quorum requirement can actually reduce turnout in a referendum.

$p(x_E, x_K, x_C)$. This is the so-called contest success function, where x_i, $i = E,K,C$, is the effort by lobby group i: the environmentalists E, the energy firm (KWR) K and the Citizens' Initiative (CI) C. The partial derivatives are $p_{xE} > 0$, p_{xK}, $p_{xC} < 0$.[11] Since both KWR and CI lobby for the same outcome, the outcome is a public good for them. A general formulation of the contest success function is given by the following logit function, axiomatized by Münster (2009) for a public good, following Skaperdas (1996) and Clark and Riis (1998) for a private good[12]:

$$p = \frac{f(x_E)}{f(x_E) + g(x_K, x_C)},$$

with f', g_{xK}, $g_{xC} > 0$. The simple Tullock (1980) function, which is often used in rent seeking analysis, is a special case of this general function with:

$$f(x_E) = x_E, \; g(x_K, x_C) = x_K + x_C,$$

All three lobbies maximize their payoffs U_i, $i = E,K,C$, given by:

$$U_E = p(x_E, x_K, x_C)\, v_E - x_E,$$

$$U_K = [1 - p(x_E, x_K, x_C)]\, v_K - x_K,$$

$$U_C = [1 - p(x_E, x_K, x_C)]\, v_C - x_C.$$

Here v_i, $i = E,K,C$, denotes lobby i's stake, i.e., the difference it makes for the lobby whether it wins or loses. When there are no binding upper limits to a lobby's effort level x_i,[13] the Nash equilibrium is determined by the first-order conditions of U_i with respect to x_i.

Since the energy firm KWR and the Citizens' Initiative are lobbying for the same outcome, the political decision is a public good for them. More specifically, as v_K will probably not be equal to v_C, it is an impure public good. Among others, Nti (1998) and Dijkstra (1999) have studied non-cooperative behaviour in contests for pure and impure public goods, respectively. Dijkstra (1999) shows that in the non-cooperative equilibrium with the simple Tullock (1980) function, only the lobby with the highest stake will be active on the pro-KWR side.

[11]Note that this formulation does not include all contests. It excludes the perfectly discriminating contest, in which the side that spends the most wins for certain (Hillman and Riley 1989). We exclude the perfectly discriminating contest here, because we do not consider it applicable to a referendum campaign.

[12]Other potential functional forms are the difference form (Hirshleifer 1989) and the relative difference form (Beviá and Corchòn 2015).

[13]See Che and Gale (1997) for an analysis of budget constraints.

However, since KWR and CI have a common interest, one might expect them to get together before the actual contest and work out a cooperation scheme. A number of specific co-operation schemes have been studied in the rent seeking literature: support (Dijkstra 1998), sharing (Loehman et al. 1996) and rewards (Baik and Kim 1997). With support, one lobby pays for part (or all) of the effort by the other lobby. Dijkstra (1998) shows that, with the simple Tullock (1980) function, the high-stake lobby will support the low-stake lobby and remain inactive itself.

3.1.3 Review of Empirical Research

In his survey on the empirical measurement of rent-seeking costs,[14] Del Rosal (2011) argues that empirical research on rent seeking has lagged behind theoretical research and that this has held back the development of the field. We might add that this is especially true for referendum campaigns.[15] Here we shall discuss research of a more qualitative nature.

Schneider and Naumann (1982) analyze the influence of interest groups on referendum voting in Switzerland. However, they only take account of the interest groups' vote recommendation, and not of additional activities.

Using data from Swiss referendums, Christin et al. (2002) show that uninformed voters tend to favour the status quo. The authors are less successful in trying to establish that uninformed voters can base their vote on cues from endorsements by political parties.

In the introduction to a special issue on the process of opinion formation and change in referendums, LeDuc (2002) distinguishes three types of dynamic. In the case of opinion formation, voters are poorly informed about the subject of the referendum. As they form their opinion, they are open to cues from various sources. The potential for volatility is very high in this case.

In the case of opinion reversal, a referendum on a reasonably well-known issue takes on a new direction in the course of the campaign. Finally, an uphill struggle occurs when a side is relatively certain of its core support, but has to reach out to other groups in order to secure victory.

In their study of the 1999 Australian republic referendum, Davidson et al. (2006) conclude that the republicans tried to argue that the change to a republic was minimal and emphasized the advantages, while the monarchists successfully argued that the change was substantial and risky.

[14]See Benito et al. (2014) and Powell (2012) for more recent empirical research and Decheneux et al. (2015) for a survey of experimental research on contests.

[15]Referendums have much in common with two-candidate elections (e.g., Hillman and Ursprung 1988; Erikson and Palfrey 2000; Ben-Bassat et al. 2015).

3.2 *The Case Study*

3.2.1 Research Questions

Our aim is to describe the Schönau contest with the notions derived from contest theory and to draw conclusions from the real world phenomena pertinent to modeling issues. Thus, we are interested not only in quantitative aspects of the contest, but also in qualitative features and in the way the lobbying interaction can be modeled. More specifically, the questions we are interested in are:

Quantitative Aspects

- How much did the lobbies spend in terms of time and money?
- How high are the lobbies' stakes?
- What are the lobbies' resource constraints? Can they spend extra effort at constant cost, as in the standard set-up, do they have hard budget constraints or no constraints at all? Specifically: Did donors and the electricity firm pay for part or all of the expenses by the environmentalists and Citizens' Initiative, respectively?
- On the eve of the referendum, how was the environmentalists' success probability assessed?

Qualitative Aspects

- What did the lobbies do: how did they try to get their arguments across, did they try to appeal to specific electoral groups?
- Were there certain electoral groups that were more inclined toward the environmentalists (EWS) or toward the energy firm (KWR)?
- In retrospect, what are the factors to which the interviewees attribute the environmentalists' victory?
- What did the lobbies know about each other's activities? Did they react to each other or, in the case of KWR and CI, fine-tune their strategies to each other?

Modeling Aspects

- What can we say about the functional form of the contest success function? How effective were the efforts by the respective groups? Which factors influence lobbying effectiveness?
- Can we summarize lobby i's efforts by one variable x_i in the contest success function, or do we have to differentiate with respect to the nature of the effort?

3.2.2 Sources

The Schönau contest was very well documented by all sides involved. We were thus able to collect not only the complete set of campaign leaflets from all three groups but also all the newspaper articles that were published concerning this case in the two local newspapers. Furthermore, in March 1999 we conducted interviews with[16]:

- Dr Michael and Mrs Ursula Sladek (environmental pressure group). Mr Sladek is the town physician and has represented the independents (FWV) in the town council since 1989;
- Mr Rolf Wetzel (environmental pressure group), policeman and manager of the environmentalists' firm EWS;
- Mrs Dagmar Zuckschwerdt (environmental pressure group), teacher in Schönau, living outside the town;
- Mr Manfred Gollin, head of the energy firm's (KWR) legal office and leader of its Schönau campaign;
- Mr Helmut Pfefferle (Citizens' Initiative), head of the CDU (conservatives) town council group from 1989-1999;
- Mr Herbert Karle (Citizens' Initiative), local CDU chairman;
- Mr Klaus Ruch (Citizens' Initiative), member of the independents (FWV);
- Mr Bernhard Seger, mayor of Schönau since 1993, CDU member.

4 Qualitative Aspects of the Contest

4.1 Composition and Organization of the Lobby Groups

4.1.1 Environmentalists (EWS)

The environmental group consists mainly of middle-class citizens not originally from Schönau. Police officer Wetzel was asked to join the group before the first referendum, because the group needed native Schönau residents like him to enhance their appeal to other native Schönau residents. The campaign team for the second referendum consisted of about ten members. The campaign was led by the Sladeks.

4.1.2 Energy Firm (KWR)

KWR is a private firm supplying energy to the Southern Black Forest (242,000 inhabitants). Its headquarters are in Rheinfelden, about an hour's drive from Schönau. In 1996, KWR's electricity mix was 60 % water power, 35 % nuclear energy and 5 % other (including Combined Heat and Power). The Schönau case was initially the responsibility of the head of the electricity customer section. Our interviewee Gollin, head of the legal office, assisted him from the beginning because of the important legal angle in the matter. Gollin took over the responsibility for the

[16]The quotes in this paper are English translations of the authorized interview reports (in German). The reports are available from the authors.

second referendum campaign, which means he was given a budget and a free hand to conduct the campaign as he saw fit.

4.1.3 Citizens' Initiative (CI)

The Citizens' Initiative was founded shortly after the town council's November 1995 decision in favour of the environmentalists. Initiated mainly by the local CDU establishment, they made an effort not to be too closely connected with the conservative CDU because they wanted to reach out to non-conservative voters as well. The social democrats (SPD) town council group leader Hitz was also on their side. Our interviewee Ruch was a member of the independents (FWV). The senior and junior managers of the plastics firm Frisetta (a.o. toothbrushes) were also associated with the Initiative. The hard core of the Initiative consisted of about ten members.

4.2 Campaign Activities by the Lobby Groups

As in any campaign, all three lobby groups organized campaign meetings and used the local media (by placing advertisements and writing letters to the editor). We now turn to the other, special activities the lobby groups undertook.

4.2.1 Environmentalists (EWS)

The environmentalists' most effective campaign weapon was house calls. They visited every household in Schönau where they considered the people to be undecided.

> "You cannot replace face-to-face conversation by any other means, because for many people it is the only form of communication they know, next to the TV." (Mr and Mrs Sladek)

Their campaign leaflets came once a week[17] for 13 weeks, and were designed in a rather simple style (black print on plain paper). The first seven leaflets consisted of one A4 sheet (double-sided), the last six leaflets consisted of a double A4 sheet. Another activity was the door-to-door distribution of jars of marmalade, produced by a local firm, with a "No" sign. On referendum Sunday, the sight of the marmalade jar on their breakfast table was supposed to remind the people to vote "No". Furthermore, the environmentalists targeted specific groups, organizing a coffee afternoon with folk music for the elderly and a rock concert for the young people on the eve of the referendum.

[17] Campaign leaflets by all three lobby groups were distributed door-to-door.

4.2.2 Energy Firm (KWR)

For the last seven weeks of the campaign, KWR established an office in Schönau for citizens with questions.[18] Like the environmentalists, KWR also delivered campaign leaflets which were published twice a week over the last 6 weeks. Each campaign leaflet was a double-sided A4 sheet, printed in blue (with the word "info" in purple) on chlorine-free bleached paper. In addition, KWR organized an equipment exhibition in the schoolyard of Schönau. There was a price tag attached to every piece of equipment in order to drive home the point that EWS would never be able to afford the equipment needed to run the electricity system.

4.2.3 Citizens' Initiative (CI)

The Citizens' Initiative, campaigning in favour of the KWR, also distributed leaflets to each Schönau household. Their six fliers came at irregular intervals and were in a simple style. Each consisted of a double-sided A4 sheet and was printed in black on coloured paper. Two leaflets were in humorous carnival rhyme. Moreover, SPD group chairman Hitz wrote an open letter in favour of the KWR, which was also delivered door-to-door. As an "answer" to the environmentalists' jars of marmalade, the pro-KWR forces distributed Frisetta toothbrushes with 10 reasons for voting "Yes". Furthermore, they organized pro-KWR advertisements by local firms and by the CDU.

4.3 Arguments from the Lobby Groups

The *environmentalists' (EWS)* main arguments were:

- we are environmentally friendly—at the same cost as before;
- the people of Schönau should take electricity supply into their own hands, instead of accepting the dictates of and paying to a big firm from outside;
- EWS attracts visitors to Schönau, directly with its energy seminars and indirectly by making Schönau a national news item.

On the other side, the *energy firm (KWR)* and the *Citizens' Initiative (CI)* argued:

- KWR is already quite environmentally friendly, with 60 % of the electricity generated by hydro power, and it is contractually tied to the present 35 % of nuclear energy;

[18]Opening hours were Mondays from 4 to 6, Wednesdays from 10 to 12 and Fridays from 3 to 6.

- EWS will not be able to operate in an economically viable way, at least not without increasing electricity prices;
- EWS does not have KWR's equipment and expertise to restore power supply in the case of failures;
- with EWS, jobs might be lost, because firms' investments are discouraged by higher electricity prices and reduced supply security;
- after the town had fully run down the electricity net in 1974, KWR bought the net and has completely modernized it since.

4.4 Strategies by the Lobby Groups

4.4.1 Environmentalists (EWS)

The campaign team kept a record of the electorate, in which they classified the voters in the categories "Yes", "No" and "undecided". They updated this classification continuously during the campaign. The goal was a total of 805 "No" (i.e., EWS) votes, which would amount to 50.1 % at a 90 % turnout. The team did not visit those classified as "Yes" votes (around 20 % at the beginning of the campaign). The "No" votes (around 20 % at the outset) were in constant need of confirmation. The campaign team tried to contact every voter through the member they thought was most sympathetic to this voter. They spent extra time trying to convince "opinion leaders" within a family.

According to Mr and Mrs Sladek, the environmentalists did not only argue factually, in their leaflets, but also emotionally, because they realized every choice has an important emotional component. In the campaign, EWS wanted to present a positive message. Zuckschwerdt and Mr and Mrs Sladek pointed out that they took no interest in the opponents' actions because replying or reacting to the opponents did not fit into their concept. They tried to ignore the opponents, but:

> "Toward the end we deviated from this strategy and reacted more in our leaflets. That was because KWR and CI produced so much sleaze, so many pathetic false claims." (M. and U. Sladek)

Knowing it was hard to dismiss the low-competence argument brought forward by KWR on a rational basis (because very few would be able to understand the calculations involved), the environmentalists framed the issue as a matter of trust. If KWR won on the basis of the wrong figures, no one would notice. But if EWS won on the basis of the wrong figures, they would be in serious (personal) trouble for failing to run the electricity system. Furthermore, the environmentalists asked: "If it is so clear that we can't run the electricity net, why is KWR trying so hard to win the referendum?" (M. and U. Sladek).

As part of their "against big business" argument, the environmentalists also took up the sheer size of KWR's lobbying effort. They argued that the rich "Goliath" KWR was trying to squash poor "David" EWS with its money, money which, in the end, they had made in Schönau. We shall refer to this effect, which played an important role in the campaign, as the "Goliath effect".

4.4.2 Energy Firm (KWR)

KWR did not have a grand strategy in the Schönau case:

> "From here we could not make a good assessment of what was going on in Schönau. The vehemence of the environmentalists surprised us again and again. We just reacted to their actions. Our people kept calling us, saying: 'They're up to something again, won't you do something?' We were under constant pressure to do more" (M. Gollin, KWR)

KWR tried to conduct a factual campaign, because as a firm and as an outsider, they could not appeal to the people's gut feelings. Only in their last leaflet, which came with a personal letter from KWR management, did they try to strike an emotional chord. The emotional side was mainly left to the Citizens' Initiative in Schönau itself which could appeal to the voters more directly. KWR reimbursed CI's material cost completely.

Gollin, head of the KWR campaign, found it hard to argue against the environmentalists' main line of approach, which he describes as: "Pro EWS means against KWR, means against nuclear energy, means pro health". His assessment of the situation was therefore:

> You can never get rid of a lie by calling it a lie. The lie is out there.

4.4.3 Citizens' Initiative (CI)

According to Ruch (CI), CI's goal was to mobilize the voters who were inclined toward the energy firm KWR but not all that interested in the matter:

> We could count on about 20 % of the votes: the loyal CDU voters and those who disliked the EWS people. At the polls, we had 48 % of the votes, so we had been able to mobilize another 28 %. The high turnout was an achievement on our part, but it was not enough. Turnout among the EWS supporters was 100 %, whereas among our supporters it was 70–80 %. We should have mobilized even more people. With a 100 % turnout (which of course is only theoretically possible) we would have won.

According to Ruch, the Citizens' Initiative always knew what the environmentalists were up to because one cannot keep that a secret in such a small town. However, he only recalls one instance where the Initiative reacted to the environmentalists' activities. After EWS had distributed marmalade to all households, the Initiative distributed Frisetta toothbrushes. This was an idea by the Frisetta manager to which Ruch and Karle had objected, arguing that such a reaction or act of retaliation would not be productive.

5 Quantitative Aspects

5.1 Stakes of the Lobby Groups

KWR was not prepared to make a statement on the size of their stake, i.e., the profit they would have made if they had been granted the energy supply concession for Schönau. We estimate this sum at approximately 8.5 m DM (i.e., the net present value of a 20-year-grant).[19]

Next to the direct financial loss, there was also the potential damage to KWR's image. The firm's image had already suffered from the conflict with the environmentalists, but the loss of image would increase further if EWS won the referendum and managed to run the electricity system. This would serve as a constant reminder that KWR had been unable to combine an environmentally-friendly and economically viable electricity supply.

Because we anticipated it would be difficult for the interviewees to give a quantitative estimate for the stakes of the environmentalists and the Citizens' Initiative, we did not try to elicit such an estimate. We did ask how much it would matter to them (in general terms) whether they won or lost, how they saw the comparative stakes of EWS and CI qualitatively and the relation between stake and lobbying effort. The interviewees agree that EWS had a higher stake than the Citizens' Initiative:

> "It would have been a catastrophe for the environmentalists if they had lost. Of course we were disappointed as well when we lost, but we could come to terms with that more easily." (H. Pfefferle, CI)

> "We had a positively formulated goal, to which we committed ourselves completely. If we had lost the referendum, that would have been the end of the Schönau Energy Initiatives. It would also have hurt us personally. We would have done a disservice to the national environmental movement, because we wanted to show that things can be done differently. There is nothing special about Schönau: if changes are possible here, they are possible anywhere. Furthermore, we would have been very annoyed that we would not be able to prove KWR's figures wrong." (M. and U. Sladek, EWS)

The interviewees also agree that the environmentalists lobbied more than the Citizens' Initiative, because they had a higher stake. This is the reason why EWS made house calls (which require a lot of time and determination) and CI did not.

[19] From the figures we have at hand, one can estimate a gross yield of 1 m DM per year. We consider the marginal personnel costs of KWR for Schönau to be negligible, since the Schönau energy consumption covered only 6 % of KWR's total output. Discounting 1 m DM gained for 20 years at 10 % gives us the stake of 8.5 m DM.

Table 1 Efforts by the lobby groups

	Time (h)	Money (DM)
EWS	2500	10–15,000
KWR	400	<30,000[a]
CI	400	5000

[a]Including all of CI's expenses

5.2 *Lobbying Effort by the Lobby Groups*

We asked the three lobby groups for a quantitative estimate of the time and money they spent on the campaign. The results are summarized in Table 1.

5.2.1 Environmentalists (EWS)

Mr and Mrs Sladek estimate the environmentalists expenses at 10–15,000 DM. The money was spent on advertisements in the local newspapers, propaganda material and copying costs, fees and travel expenses for invited speakers and artists and the jars of marmalade. The latter was the largest single item, costing a few thousand DM.

The environmentalists simply spent what they thought was needed. They were planning to try and cover their expenses afterward through donations. If the donations had fallen short of the expenses, they would have paid the difference from their own pockets.

Mr and Mrs Sladek were only able to give a very rough estimate of the time spent on the campaign. The estimate amounts to 2500 h, consisting of the following components:

- For every brochure, around 10 people got together for 4 h to discuss the contents. The result was then processed by Mrs Sladek (5 h) and a layout expert (15 h). Distributing the brochures to every household in Schönau, which was mainly a task for the members' children,[20] took another 5 h. This amounts to 65 h per brochure, which multiplied by 13 brochures yields 850 h.
- In the first phase of the campaign (the first 10 weeks), the 10 active members spent about one hour a day talking to Schönau inhabitants about the concession grant.[21] This amounts to 70 h a week, 700 h altogether.
- In the hot phase of the campaign (the last three weeks), police officer Wetzel took two and physician Sladek three weeks off in order to campaign full time. At a rate of 60 h a week, this amounts to 300 h.

[20]We realize it is questionable whether time spent by the members' children should count as time spent by the environmentalists.

[21]According to the Sladeks, a house call could take up to 1½ hours. Wetzel took between 10 min and 2 h, on average at least 20 min, whereas Zuckschwerdt needed 2 to 3 h.

– During the last three weeks, the other eight members of the campaign team spent about four hours a day campaigning. This amounts to $8 \times 4 \times 20 = 640$ h.

According to the EWS interviewees, the team spent all their spare time campaigning:

> "We gave it all for the victory. If it hadn't been enough, at least we would not have ourselves to blame. If people don't want change, one has to accept that." (M. and U. Sladek)

5.2.2 Energy Firm (KWR)

For the referendum campaign, Gollin requested and received a budget of 30,000 DM. He did not exhaust this budget. The brochures were the most expensive item: the fee for the designer, the printing and the distribution.

Gollin reconstructs KWR's input of time as follows:

– Gollin visited Schönau around 8 times, and had perhaps two meetings with colleagues in the nearby town of Zell. At 4 h per visit, this amounts to 40 h.
– Gollin spent 4 h on every brochure, which for 12 brochures amounts to 48 h.
– The head of KWR's electricity customer section always accompanied Gollin to Schönau, and all in all must have spent the same amount of time: 90 h.
– The lady from the public relations office spent around half as much time as Gollin: 45 h.
– The equipment exhibition in Schönau: one day for a number of KWR employees.
– The Schönau campaign office was open for a total of 50 h. Per employee present, travelling time would be another 50 h. During opening hours, one or two KWR employees were present. Taking one and a half as an average, this amounts to 150 h.

Thus, total estimated time input from KWR is about 400 h, 180 of them at the executive level and 220 at a lower level.

According to Gollin, there was a heated debate within KWR whether they should not do more. Gollin named two reasons for the modest size of KWR's input. The first is that KWR did not want to leave the impression that the rich "Goliath" KWR was steamrollering poor "David" EWS with its money. Secondly, Gollin realized that the people would be motivated by their gut feelings. Thus, spending more money on glossy brochures or posters would not have brought much:

> An expensive publicity campaign makes sense to introduce a new product or a brand name. But that was not the issue here. Every voter in Schönau knew KWR and EWS. It is not possible for an outside actor like us to pull anyone over to our side with a publicity campaign. Especially not in Schönau where the positions had hardened so much that some people did not even talk to each other anymore.

5.2.3 Citizens' Initiative (CI)

According to Karle, the CI treasurer, the Citizens' Initiative spent around 5000 DM. KWR paid for all the expenses. According to Karle, there was no budget constraint from KWR. The Initiative could have spent much more, for instance on a professional brochure, and KWR would still have paid for everything. But the initiative did not want to spend on such things, because "we did not want to flood people with campaign material" (Karle).

Ruch estimates his own time input at at least 10 h a week for 10 weeks. He estimates the average input of the other 10 members at 30 h at the most. This amounts to a total of 400 h.

5.3 Lobbying Productivity

We tried to elicit estimates of marginal lobbying productivity from the interviewees, with questions like: "Suppose the energy firm KWR had spent another 100 DM (be it in promotion material or in time) on the campaign. How much would the Citizens' Initiative have had to spend to accomplish the same effect? How much would the environmentalists have had to spend to offset the effect?" Only Gollin, head of the energy firm's campaign, was able to give a quantitative estimate: he considered 100 DM from the Citizens' Initiative to be at least as effective as 500 DM from KWR.

All of the interviewees agreed that KWR's campaign productivity was lowest, because the firm could not appeal to the voters as personally and emotionally as the environmentalists or the Citizens' Initiative. As Gollin (KWR) and Pfefferle (CI) pointed out, the energy firm would not have been able to collect signatures for the referendum in the first place. Another handicap for KWR, mentioned by Mr and Mrs Sladek (EWS), was that they came from outside. In the ultimate phase of the campaign, outside support is futile. Finally, the interviewees agree that the environmentalists' attempts to turn extra lobbying effort by KWR against them, referred to in Sect. 4.4 as the "Goliath effect", (possibly) constituted a problem for KWR.

Our interviewees disagree on whether the environmentalists or the local pro-KWR forces achieved higher campaign productivity. The heads of the environmentalists' campaign, Mr and Mrs Sladek, estimate that their campaign productivity was lower than CI's, as EWS started the campaign from a very bad position (see Sect. 6.2.6). The environmentalists compensated this disadvantage with high lobbying input, whereas the Citizens' Initiative managed to achieve considerable results with relatively little effort. According to Ruch (CI), however, the environmentalists' campaign was more productive, because they targeted specific voter groups, like the elderly and the youth.

5.4 Success Estimates on the Eve of the Referendum

On the eve of the referendum, both sides felt confident. Only Gollin, campaign manager for the energy firm KWR, says he was too far away from Schönau for a reliable assessment of the odds:

> The people from the Citizens' Initiative kept telling us: 'We are going to win. Everyone we meet is against the Sladeks.' But I know from my own experience how that works: You only talk to the like-minded. The opponents don't talk to you or don't speak up.

This quote also provides us with one of the reasons why both sides felt confident. However, there were also concerns on both sides. Ruch (CI) was somewhat worried about the rock concert organized by EWS on Saturday night. He realized there were still a considerable number of votes to be gathered from the youth. Wetzel (EWS) was worried about the high turnout[22]:

> We had counted on 600 sure votes, which would have been sufficient at 65 % turnout of an 1800 voter electorate. But as turnout rose above 80 %, we were not sure whether we had been able to mobilize enough voters on our side.

6 The Vote

6.1 Voter Groups

The interviews yield a consistent picture on the issue of which voter groups were more inclined to vote one way or the other. Tending more toward the environmentalists were:

– the older people, who had great confidence in their physician Dr. Sladek. Furthermore, they were receptive to the argument that a vote for EWS was a vote against atomic energy and for a safer world for their grandchildren. Winning over the older people was an important coup for EWS, because they traditionally vote conservative;
– the youth, who were against atomic energy. They wanted to rebel against "big bad business" embodied by KWR. Finally, Sladek's boisterous, easy-going attitude appealed particularly to them;
– the non-natives who, having already moved once, "were more open to new things" (Zuckschwerdt, EWS).

Tending more toward the energy firm KWR were:

[22]Note that Ruch (CI) also considered a high turnout to be working in favour of the energy firm KWR (see Sect. 4.3).

Table 2 Breakdown of the referendum vote

	District 1	District 2	Postal votes	Total
Allowed to vote	797	987		1784
Valid votes	540	720	233	1493
Yes (KWR)	273 (50.6 %)	348 (48.3 %)	90 (38.6 %)	711 (47.6 %)
No (EWS)	267 (49.4 %)	372 (51.7 %)	143 (61.4 %)	782 (52.4 %)

Note: Turnout was 84.3 %. Of 1504 cast votes, 11 were not valid

– the native residents of Schönau. Their attitude is best summed up by Pfefferle's (CDU) statement in the November 1995 town council meeting:

> In electricity matters, in environmental matters, in economic matters, in service security matters, we have no problems in Schönau. And when there are no problems, why change anything?

– the middle-aged;
– the lower-education, lower-income groups, especially those working with local firms. They were afraid that with EWS, electricity prices would rise and employment would be harmed, because firms might leave or not expand in Schönau.

The breakdown of the referendum vote (Table 2) corroborates the impressions from the interviewees. EWS did best with the postal votes. These are mostly elderly people, or young people living outside Schönau. Electoral district 1 is the lower-income district of the two. In this district, KWR actually defeated EWS.

6.2 Success Factors

In this subsection, we look at the factors which influenced, and maybe even determined, the outcome of the referendum. The factors are ordered in roughly descending order of importance, as our interviewees saw it. In the concluding Sect. 6.2.6, we present our own assessment of the success factors.

6.2.1 Dr. Sladek

All the other interviewees see Dr. Sladek as the driving force behind the environmentalists' strategy of confrontation with the energy firm KWR. They also ascribe the ultimate success of this strategy largely to his involvement. Especially the non-EWS interviewees emphasize Sladek's role in the referendum campaign. The importance of Sladek's involvement consists of the following elements:

– As a town physician, Sladek knew and had access to many people, and people, especially the elderly, had great confidence in him;

– his personality, described as charismatic, inspiring and congenial;
– his commitment, which prompted him to invest a lot of time in house calls.

On the other hand, Gollin (KWR) and Ruch (CI) also mention that Sladek's public image antagonized many people "who don't like his way of walking through the streets like a guru" (M. Gollin).

6.2.2 Lobbying Quality and Quantity

The interviewees agree that both the quantity and the quality of the environmentalists' lobbying input was highest, and that this contributed to their victory. As far as lobbying quantity mattered, it was the time spent, not the money:

> "Money hardly played a role in the campaign. None of the groups tried to win votes by spending lots of money. Nor would it have worked that way." (K. Ruch, CI)

The main difference in lobbying input was that the environmentalists made house calls and the Citizens' Initiative did not. Pfefferle, Karle and Ruch (all members of the Citizens' Initiative) realize they might have won if they had also made house calls. But they admit the matter was just not important enough for them to spend so much time and effort on house calls, even if it would have guaranteed the victory.

The pro-KWR side is divided on the question whether the energy firm itself and/or the supporting Citizens' Initiative should have done more. While recognizing a possible "Goliath" effect (Sect. 4.4), Ruch (CI) still believes that KWR did not do enough:

> KWR could have done more, and it wouldn't necessarily have cost that much. A bit more local presence, for instance with radio spots, in newspapers or with posters, would have done a lot of good. If KWR had made an ultimate effort in the final days, we would have won. A few days before the vote, we warned KWR that they should do more. But they must have thought that victory was already theirs.

As we have seen in Sect. 5.2, there was also a call from within the energy firm for more involvement, but Gollin, head of the KWR campaign, successfully defended his modest spending outlay. Gollin himself entertains the thought that KWR should perhaps have tried to activate the pro-KWR forces in Schönau at an earlier stage. However this strategy has its limitations:

> As a firm, you can't just go to someone you don't know personally and say: 'Do something for us.' All you can do is refund the expenses for such an initiative. You can't demand from them to put their time into it.

The Sladeks (EWS) also suggest that, since more spending by KWR would not have been effective, the energy firm should have pushed the Citizens' Initiative more. Pfefferle (CI) tends to agree, but Ruch (CI) argues against it:

> It was our task to motivate our supporters, and you can only do that in the last couple of weeks before the vote.

6.2.3 Campaign Quality, Quality of Arguments, Credibility

Zuckschwerdt and the Sladeks (all EWS) name the quality of the environmentalists' campaign, the quality of their arguments and their credibility as important success factors. In the previous years, the environmentalists' activities (recorded in Sect. 2) had earned them credibility and respect, even from those who were initially suspicious of the "eco-freaks". Ruch (CI) also mentions that the environmentalists' strategy of inviting renowned experts and TV personalities to Schönau boosted their credibility.

In the campaign, EWS tried to remain factual and not to return the innuendos from the other side in kind, and this paid off.

> "The other side could not come up with anything concrete and just went berserk." (D. Zuckschwerdt, EWS)

According to Gollin (KWR), it was just the other way around:

> Dr. Sladek tapped into basic fears, which was hard to counter with the more factual information from KWR.

Ruch (CI), however, agrees with the environmentalists' view:

> EWS had the better arguments: pro-environment, pro-future. All we could do was scare the people a bit. The substance of our position was much weaker, because we were merely against something. These days, the progressive have an advantage over those who want to take it slow. We only had a few trumps in the campaign: first Frisetta's warning of employment losses and secondly the conservative attitude of the people in Schönau. But the people were just not conservative enough.

6.2.4 The Mayor

Although personally in favour of KWR, mayor Seger kept the promise made in his 1993 mayoral campaign that he would not speak out officially in favour of either position. Not only did he abstain from the November 1995 town council vote, he also refrained from endorsing either side during the referendum campaign. Karle and Pfefferle (CI), CDU members like Seger, are convinced that KWR would have won if Seger had come out in favour of the energy firm. Mayor Seger disagrees:

> I think 80 % of the voters knew my personal stance in the matter and how I would vote. I don't believe the outcome would have changed if I had spoken out publicly as a mayor. Voters are not that manipulable.

6.2.5 The Town Council Vote

A corrective referendum is a chance to repeal a prior town council vote. One side lobbies against the town council vote, the other side lobbies in favour of it. It is an interesting question which of these lobbies has a better starting position. This is particularly interesting when modeling the multi-stage game introduced in

Sect. 3.1.1. When it is an advantage to lobby in favour of the town council decision, this is a reason for trying to win the majority in the town council.

The interviewees in Schönau are experienced in this matter, because the town council majority was reversed between its July 1991 and its November 1996 decisions. The environmentalists campaigned against a town council decision at the first referendum, but in favour at the second referendum. For the energy firm KWR and its allies, the opposite applies.

According to Ruch and Pfefferle (both CI), it is an advantage to lobby against a town council decision, because of the "protest potential" of people who like to vote against. Wetzel and the Sladeks (all EWS), on the other hand, argue that the town council vote for EWS worked in their favour at the second referendum. In their view, a town council decision is an important argument in convincing predominantly conservative and authority-abiding voters.

However, although the vote was in their favour, the environmentalists saw themselves in a bad starting position immediately after the town council meeting. At the meeting, both EWS and KWR had presented their offers. While EWS had made a factual presentation, KWR seized the opportunity for an aggressive attack, calling into question the viability of EWS's offer.

> "After this meeting, we first had to reassure our own supporters that our project was not only ecologically worthwhile, but also economically viable." (U. and M. Sladek, EWS)

6.2.6 Assessment of the Success Factors

To conclude the analysis of the success factors, we present our own assessment of EWS's success. We attach great importance to the role of Dr Sladek, because of his charisma and his position within the town as a physician. Furthermore, having worked for years on the subject, the environmentalists not only had a high stake in the outcome. They had also built up team spirit, credibility with the voters and experience in persuading people.

Another vital difference between the environmentalists and the energy firm is the availability of spare time by volunteers. This lobbying input was far more productive than money, the other input. The environmentalists put all their spare time into the campaign. KWR, as a firm, did not have direct access to spare time. They had the local Citizens' Initiative working for them, but its members were not motivated enough to spend large amounts of spare time campaigning. All that KWR could do (and did) was refund the Initiative's material expenses. Paying for their time was out of the question, since this would have destroyed the very nature of the input.

The environmentalists' campaign was not only labour-intensive, but also of high quality.[23] For each voting group (the youth, the elderly, the parents, etc.) an individual campaign approach was chosen. Dr Sladek as their figurehead, the

[23] As in the 1999 Australian republic referendum (Davidson et al. 2006), the pro-change side made a positive case for change, while the pro-status quo side emphasized the risks of change.

support of the elderly and the town council vote in their favour functioned as cues for uninformed voters (Christin et al. 2002; LeDuc 2002).

On the other side, the energy firm KWR cleverly used the town council meeting to spread doubts about the economic viability of the environmentalists' bid. KWR's tactic of addressing the voters over the heads of the town council may be seen as their only truly effective action. KWR had the advantage of representing the status quo (Christin et al. 2002), but it seems that they underestimated the environmentalists throughout. In the campaign, the Goliath effect kept them from spending as much as they would have wanted to. We should note here that the Goliath effect only exists because the environmentalists managed to turn the size of KWR's lobbying effort against them. This effect is therefore another token of the quality of the environmentalists' campaign. Finally, the Citizens' Initiative, while necessary for KWR to stand a chance at all, did not have the motivation, coherence and skills to make the difference.

7 Modeling Issues

In this section, we shall analyze the referendum campaign in Schönau from the perspective of the theory of the rent seeking contest. Taking the size (Table 1) and the nature of the lobbies' efforts as our point of departure, we shall see how this case study can enrich the theory.

The first point to note is the low level of KWR's lobbying input relative to its stake, tentatively calculated in Sect. 5.1 as 8.5 m DM. The apparently low level of rent seeking effort is a general phenomenon that has puzzled researchers (Tullock 1997). Two main reasons can be given for KWR's low effort level.

The first reason is that we have to distinguish between the lobbying inputs of money (be it spent on material, wages or external services) and spare time. Spare time was by far the most productive input. As a firm, KWR did not have direct access to spare time, and thus its money input was of limited use.

Further evidence of the superior importance of spare time relative to money comes from observing the input choices of the environmentalists EWS and the pro-KWR Citizens' Initiative. Both had practically zero opportunity cost of spending money. The Citizens' Initiative's expenditures were covered by KWR, and EWS planned to acquire funds from sponsors. But however cheap their money was, EWS and CI did not spend large sums. This indicates that the lobbying productivity of money is low. Spare time, however, was a precious lobbying input. The environmentalists' stake was so high that they spent all their spare time campaigning. The Citizens' Initiative did not spend nearly as much time, because their stake was much lower.

Rent seeking theory has, until now, mainly regarded lobbying effort as a homogeneous good. In order to model complementarity between time and money, one can define x_i as lobby i's effective or aggregate lobbying output. This output is then a composite of time $x_i{}^T$ and money $x_i{}^M$. Complementarity implies: [24]

$$\frac{\partial^2 x_i\left(x_i^T, x_i^M\right)}{\partial x_i^T \partial x_i^M} > 0.$$

We could go even further by modeling extra money, unmatched by extra time, as completely useless. Then, with suitable choice of units, time and money would be complementary for $x_i{}^T > x_i{}^M$, but $x_i = x_i{}^T$ for $x_i{}^T \leq x_i{}^M$. In our application, EWS and CI had spare time and money available as substitutes, while KWR could only spend money (either on expenses or personnel).

Epstein and Hefeker (2003) model a contest between two parties who have two inputs, or instruments, available for lobbying. The instruments are complementary, with one input being essential and the other optional. However, in our application it seems more appropriate to model both spare time and money essential to EWS and CI. Arbatskaya and Mialon (2010) axiomatize the contest success function for a multi-activity contest between two players.

Schoonbeek (2007) models a contest between two parties who can either compete themselves, using one instrument, or hire a delegate who can use two instruments, as in Epstein and Hefeker (2003). In our application, CI could be seen as KWR's delegate. However, there are several differences between the contest in Schoonbeek (2007) model. First, CI had a stake in the contest, whereas Schoonbeek's (2007) delegates do not have any stake. Secondly, KWR paid all of CI's monetary expenses, whereas in Schoonbeek's (2007) model a player offers the delegate a contingent fee. Thirdly, in Schoonbeek's (2007) model a delegating player does not compete, but KWR itself also competed in Schönau. Finally, EWR had spare time and money at its disposal, whereas in Schoonbeek's (2007) model it would only have one input.

The second reason for the small lobbying effort by the electricity firm KWR is that they experienced the Goliath effect: the size of their lobbying effort worked against them. It would be interesting to model the possibility that success probability becomes declining in own effort. One can introduce the Goliath effect for KWR in the contest success function from Sect. 3.1.2, while retaining the simple Tullock (1980) form for the environmentalists and the Citizens' Initiative's efforts by setting:

$$f_E\left(x_E\right) = x_E,$$

$$g\left(x_K, x_C\right) = g_K\left(x_K\right) + x_C,\ \ g_K\left(x_K\right) = \frac{\left(x_K + 1\right)^r}{r} - 1,$$

[24] As long as it is not of the strict Leontief type.

with $0 < r < 1$, so that $g_K'(0) = 1$ and $g_K'' < 0$. A further distinction can be made between a weak Goliath effect, where g_K is monotonically increasing in x_K, and a strong Goliath effect, where g_K is decreasing for large x_K.[25]

When modeling the Goliath effect, David should also be taken into account. EWS could only play on the Goliath effect because they had limited means themselves. Thus, David players should be equipped with a budget constraint.

Another point to note with respect to lobbying activity is that KWR paid for all the expenses run up by CI. Thus, the support mechanism analyzed by Dijkstra (1998) was at work, but it was differentiated with respect to lobbying inputs. There was full support for one lobbying input (money), but no support for the other (time). This is arguably the maximum support that KWR had to offer. Had KWR paid CI for their time, they would have destroyed the very nature of this highly productive input. Full support of money input can therefore be regarded as a corner solution. Lobbying activity by CI was very important for KWR, so they supported the Initiative as much as they could.

Although KWR and CI were on the same side and KWR supported CI, both KWR and CI had positive lobbying efforts. This limits the class of contest success functions applicable to the Schönau contest. We have noted in Sect. 3.1.2 that with the simple Tullock (1980) contest success function, KWR remains inactive in the contest when they have supported CI.

The difference between the simple Tullock (1980) function and the referendum campaign is that in the Tullock function, lobbying efforts by KWR and CI are perfect substitutes: they are added up and treated as if they came from one agent. However, lobbying effort by KWR and CI was in fact complementary. KWR supplied the factual information and conveyed the image of experience and reliability, whereas CI made an emotional appeal to the voters.

Finally, we saw that the environmentalists classified voters into "Yes", "No" and "undecided", and targeted specific voter groups. The Citizens' Initiative saw it as their main task to mobilize lukewarm support. This points to the need for including heterogeneity of voters in a contest model for a referendum campaign. For instance, as is common practice in voting models, voters can be placed in a continuum between the "Yes" and "No" alternatives. The closer a voter's position to the lobby's alternative, the easier it will be for the lobby to obtain this vote. Another element to be included is that a lobby should not only get a voter to prefer its option, it should also increase this preference to a level where the voter will actually go and vote.

[25] Note that the CSF with the Goliath effect is not homogeneous of degree zero: A doubling of everyone's effort would not leave p unchanged, indeed it would reduce p. Homogeneity of degree zero is often seen as a desirable property of a CSF (e.g., Skaperdas 1996; Münster 2009).

8 Conclusion

After the introduction of a simple contest success function by Tullock (1980), the theory of the rent seeking contest has expanded in many ways. By far the largest proportion of the contributions are hardly motivated by empirical considerations. This can be partly explained by the difficulty of performing empirical research into rent seeking activities. However, one promising field of empirical research has been overlooked so far. These are referendum campaigns.

In this paper, we have analyzed a referendum campaign between environmentalists and an electricity firm with its allies in the small German town of Schönau. This campaign can fruitfully be analyzed as a contest. The players interviewed present a consistent picture of the relations between efforts, stakes and success probability. The contest did have a number of special features that have not been modeled before, but this is more a reason for modeling these features as well rather than rejecting the applicability of the contest model.

The issues worth modeling are first, that lobbying inputs should be diversified into money and spare time. Time has emerged as the more productive and the key strategic variable. We are convinced that this is a central feature of many political campaigns; especially when they take place in a regional or local setting. But even national election campaigns rely heavily on the time input of (party) members, who voluntarily organize campaign booths on the marketplace etc. Secondly, the electricity firm seems to have suffered from a "Goliath effect", rendering its efforts progressively less productive. Thirdly, we found evidence of complementarity between efforts by lobby groups on the same side. Finally, in a rent seeking contest for a referendum, voter heterogeneity and turnout should be taken into account.

References

Arbatskaya, M., & Mialon, H. M. (2010). Multi-activity contests. *Economic Theory, 43*, 23–43.

Baik, K.-H. (1994). Effort levels in contests with two asymmetric players. *Southern Economic Journal, 61*, 367–378.

Baik, K.-H., & Kim, I.-G. (1997). Delegation in contests. *European Journal of Political Economy, 13*, 281–298.

Ben-Bassat, A., Dahan, M., & Klor, E. F. (2015). Does campaign spending affect electoral outcomes? *Electoral Studies, 40*, 102–114.

Benito, B., Bastida, F., Ríos, A.-M., & Vicente, C. (2014). The causes of legal rents extraction: Evidence from Spanish municipalities. *Public Choice, 161*, 367–383.

Beviá, C., & Corchòn, L. C. (2015). Relative difference contest success function. *Theory and Decision, 78*, 377–398.

Che, Y.-K., & Gale, I. (1997). Rent dissipation when rent seekers are budget constrained. *Public Choice, 92*, 109–126.

Christin, T., Hug, S., & Sciarini, P. (2002). Interests and information in referendum voting: An analysis of Swiss voters. *European Journal of Political Research, 41*, 759–776.

Clark, D. J., & Riis, C. (1998). Contest success functions: An extension. *Economic Theory, 11*, 201–204.

Congleton, R. D., Hillman, A. L., & Konrad, K. A. (2008). *40 years of research on rent seeking*. New York: Springer.

Cornes, R., & Hartley, R. (2005). Asymmetric contests with general technologies. *Economic Theory, 26*, 923–946.

Cornes, R., & Hartley, R. (2007). Aggregative public good games. *Journal of Public Economic Theory, 9*, 201–219.

Cornes, R., & Hartley, R. (2012). Risk aversion in symmetric and asymmetric contests. *Economic Theory, 51*, 247–275.

Davidson, S., Fry, T. R. L., & Jarvis, K. (2006). Direct democracy in Australia: Voter behaviour in the choice between constitutional monarchy and a republic. *European Journal of Political Economy, 22*, 862–873.

Decheneux, E., Kovenock, D., & Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics, 18*, 609–669.

Del Rosal, I. (2011). The empirical measurement of rent-seeking costs. *Journal of Economic Surveys, 25*, 298–325.

Dijkstra, B. R. (1998). Cooperation by way of support in a rent seeking contest for a public good. *European Journal of Political Economy, 14*, 703–725.

Dijkstra, B. R. (1999). *The political economy of environmental policy: A public choice approach to market instruments*. Cheltenham (UK): Edward Elgar.

Epstein, G. S., & Hefeker, C. (2003). Lobbying contests with alternative instruments. *Economics of Governance, 4*, 81–89.

Erikson, R. S., & Palfrey, T. R. (2000). Equilibria in campaign spending games: Theory and data. *American Political Science Review, 94*, 595–609.

Graichen, P. R. (2003). *Kommunale Energiepolitiek und die Umweltbewegung: Eine Public-Choice-Analyse der "Stromrebellen" von Schönau*. Frankfurt/New York: Campus Verlag.

Graichen, P. R., Requate, T., & Dijkstra, B. R. (2001). How to win the political contest: A monopolist vs environmentalists. *Public Choice, 108*, 273–293.

Herrera, H., & Mattozzi, A. (2010). Quorum and turnout in referenda. *Journal of the European Economic Association, 8*, 838–871.

Hillman, A. L. (1989). *The political economy of protection*. London: Harwood Academic Publishers.

Hillman, A. L., & Riley, J. G. (1989). Politically contestable rents and transfers. *Economics and Politics, 1*, 17–37.

Hillman, A. L., & Ursprung, H. W. (1988). Domestic politics, foreign interests, and international trade policy. *American Economic Review, 78*, 729–745.

Hirshleifer, J. (1989). Conflict and rent-seeking success functions: Ratio versus difference models of relative success. *Public Choice, 63*, 101–112.

LeDuc, L. (2002). Opinion change and voting behaviour in referendums. *European Journal of Political Research, 41*, 711–732.

Liston-Heyes, C. (2001). Setting the stakes in an environmental contest. *Journal of Environmental Economics and Management, 41*, 1–12.

Loehman, E., Quesnel, F. M., & Babb, E. M. (1996). Free-rider effects in rent-seeking groups competing for public goods. *Public Choice, 86*, 35–61.

Long, N. V. (2013). The theory of contests: A unified model and review of the literature. *European Journal of Political Economy, 32*, 161–181.

Münster, J. (2009). Group contest success functions. *Economic Theory, 41*, 345–357.

Nti, K. O. (1998). Effort and performance in group contests. *European Journal of Political Economy, 14*, 769–781.

Nti, K. O. (1999). Rent-seeking with asymmetric valuations. *Public Choice, 98*, 415–430.

Powell, B. (2012). Coyote ugly: The deadweight cost of rent seeking for immigration policy. *Public Choice, 150*, 195–208.

Schneider, F., & Naumann, J. (1982). Interest groups in democracies-How influential are they? An empirical examination for Switzerland. *Public Choice, 38*, 281–303.

Schoonbeek, L. (2007). Delegation with multiple instruments in a rent-seeking contest. *Public Choice, 131*, 453–464.

Skaperdas, S. (1996). Contest success functions. *Economic Theory, 7*, 283–290.

Tullock, G. (1967). The welfare cost of tariffs, monopoly and theft. *Western Economic Journal, 5*, 224–232.

Tullock, G. (1980). Efficient rent seeking. In J. M. Buchanan, R. D. Tollison, & G. Tullock (Eds.), *Toward a theory of the rent-seeking society* (pp. 97–112). College Station: Texas A&M Press.

Tullock, G. (1997). Where is the rectangle? *Public Choice, 91*, 149–159.