

Susumu Egashira  
Masanori Taishido  
D. Wade Hands  
Uskali Mäki *Editors*

# A Genealogy of Self-Interest in Economics

 Springer

# A Genealogy of Self-Interest in Economics

Susumu Egashira · Masanori Taishido ·  
D. Wade Hands · Uskali Mäki  
Editors

# A Genealogy of Self-Interest in Economics

 Springer

*Editors*

Susumu Egashira  
Department of Economics  
Otaru University of Commerce  
Otaru-shi, Hokkaido, Japan

Masanori Taishido  
Faculty of Economics  
Toyo University  
Tokyo, Japan

D. Wade Hands  
Department of Economics  
University of Puget Sound  
Tacoma, WA, USA

Uskali Mäki  
Political and Economic Studies  
University of Helsinki  
Helsinki, Finland

ISBN 978-981-15-9394-9      ISBN 978-981-15-9395-6 (eBook)  
<https://doi.org/10.1007/978-981-15-9395-6>

© Springer Nature Singapore Pte Ltd. 2021

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Contents

|   |     |
|---|-----|
| <b>Introduction to “A Genealogy of Self-interest in Economics”</b> . . . . .  | 1   |
| D. Wade Hands   |     |
| <b>Self-Interest in Eighteenth and Nineteenth Century Economics</b>   |     |
| <b>Self-interest and French ‘Philosophie économique’ 1695–1830</b> . . . . .  | 11  |
| Gilbert Faccarello and Philippe Steiner   |     |
| <b>Self-interest in the Thought of Adam Ferguson</b> . . . . .  | 31  |
| Craig Smith   |     |
| <b>Adam Smith on Self-interest</b> . . . . .  | 47  |
| Shinji Nohara   |     |
| <b>Bentham on Self-interest: Institutional Control of Self-interest</b> . . . . .                                   | 61  |
| Hiroaki Itai  |     |
| <b>John Stuart Mill on Self-interest: Focusing on His Political Economy<br/>and the <i>Principles</i></b> . . . . . | 85  |
| Yoshifumi Ozawa   |     |
| <b>Léon Walras on Human Nature and His Social Reform Plan</b> . . . . .   | 107 |
| Satoshi Takahashi   |     |
| <b>P. H. Wicksteed on Self-interest: Resource Allocation<br/>and Social Inclusion</b> . . . . .                     | 125 |
| Yoshio Inoue  |     |
| <b>Self-Interest in Twentieth Century Economics</b>   |     |
| <b>Otto Neurath’s Theory of Felicitology and the Will<br/>to Socialization</b> . . . . .                            | 147 |
| Manabu Kuwata   |     |

|   |     |
|---|-----|
| <b>Selfish, Therefore Reciprocal: The Second Marginal Revolution of Mises</b> .....   | 167 |
| Akihiko Murai   |     |
| <b>Frank Knight on Self-interest</b> .....  | 189 |
| Masanobu Sato   |     |
| <b>Karl Polanyi’s Motive of Economy and Institution</b> .....   | 203 |
| Takato Kasai  |     |
| <b>Contributions of Science of Free Will to Neuroeconomics and Quantum Decision Theory</b> .....  | 221 |
| Taiki Takahashi   |     |
| <b>An Institution to Reconcile Self-interests: Hayek on the Notion of the Economic Agent and the Establishment of Appropriate Rules</b> ..... | 235 |
| Masanori Taishido   |     |
| <b>Contemporary Topics in Self-Interest and Economics</b>   |     |
| <b>Creating an Algorithm Based on the Theory of Moral Sentiments</b> .....  | 261 |
| Susumu Egashira   |     |
| <b>The “Self” and the “Others”: From Game Theory to Behavioral and Neuroeconomics</b> .....   | 273 |
| Christian Schmidt   |     |
| <b>Why Is Behavioral Game Theory a Game for Economists? The Concept of Beliefs in Equilibrium</b> .....                                       | 289 |
| Michiru Nagatsu and Chiara Lisciandra   |     |
| <b><i>Homo Economicus</i> Under Multiple Pressures</b> .....  | 309 |
| Uskali Mäki   |     |

# Introduction to “A Genealogy of Self-interest in Economics”



## D. Wade Hands

It can be argued that economics has always been concerned with self-interest. The most quoted passage from Adam Smith’s *Wealth of Nations* has been his remarks about the *invisible hand*: a good social result that emerges from self-interested behavior.

... by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public good (Smith 1776, B4/C2, p. 423).

Although this quote has been subject to various interpretations and reinterpretations over the years, the most straightforward reading is that self-interest will (or at least can) serve the public good more effectively than actions motivated primarily by social or public concern. Of course the good social result requires the right institutional conditions—the rule of law, competitive markets, etc.—but assuming those conditions are met, then the invisible hand suggests that people being self-interested is a socially good thing. And even earlier in the *Wealth of Nations* there is another, almost as frequently cited, remark concerning self-interest:

It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages (Smith 1776, B1/C2, p. 11).

Here the role of self-interest is more about a putative fact of individual psychology: what one should appeal to in order to motivate another person to take a particular

---

D. W. Hands (✉)

Department of Economics, University of Puget Sound, Tacoma, WA 98416, USA

e-mail: [hands@pugetsound.edu](mailto:hands@pugetsound.edu)

© Springer Nature Singapore Pte Ltd. 2021

S. Egashira et al. (eds.), *A Genealogy of Self-Interest in Economics*,

[https://doi.org/10.1007/978-981-15-9395-6\\_1](https://doi.org/10.1007/978-981-15-9395-6_1)

action. The invisible hand passage is more about how to bring about certain desirable social outcomes, whereas the second passage is more about how to motivate individual behavior, but both clearly emphasize the role of self-interest in economic life.

This commitment to self-interest by economists not only started early in the discipline's history, it has: persisted over the next two hundred-plus years of economic theorizing, continued to condition the way that economists characterize individual and institutional decision-making, and influenced the way that economists think about public policy questions. To some degree economics *is* the social science that studies the consequences of individual self-interested action. Sometimes the relevant consequences are the intentional outcomes of self-interested actions, and sometimes the important economic outcomes are the unintended consequences of such actions. These unintended consequences can of course be good (such as the invisible hand) or bad (such as a prisoner's dilemma). This commitment to self-interested motivations and consequences by economists has often polarized discussions about the discipline of economics. Critics often complain that economists are so committed to self-interest that they fail to recognize that humans have social and moral values that influence their actions, even economic actions, in important ways. It is also argued that by only considering self-interested behavior economists have a dangerously narrow take on social theorizing, social phenomenon, and social policy. On the other hand, many economists seem to be just as polarized in the opposite direction. Some defend economics by arguing that all human action—or at least the vast majority of such action—is primarily self-interested and thus economics is, and should be, the master social science. While others argue for an intellectual division of labor; economics is the science of self-interested action, other social sciences can define themselves in any way they see fit.

One of the main themes of this volume is to challenge this polarized vision of the relationship between economics and self-interest. The goal is not to take sides with either of these positions—although individual authors may have their own views on the matter—but rather to demonstrate historically, and to some degree philosophically, that the whole debate is based on an erroneous presupposition about the relationship between economics and self-interest. The erroneous presupposition is that there has been a fairly simple, consistent, and tension-free relationship between economics and self-interest during the last two hundred-plus years. The fact is that self-interest has played a quite complex and often tension-ridden role in the history of economic thought; it has served many different functions and has been instantiated in many different ways (between schools of thought certainly, but often among economists of the same school, and even in certain cases within the work of a single economist). As a result, the history of economic thought is replete with various concepts of self-interest working in a number of diverse ways. The complexity of self-interest in economics is exhibited to some degree in virtually every chapter of the volume, but I will leave it to the authors to present their own arguments. My goal in this introduction is to help motivate interest in an in-depth historical discussion of the self-interest/economics relationship. To that end I will provide three examples: one



comes from classical economics, one comes from the twentieth century economics, and one is a fairly recent development.

Consider for example the classical economics of Ricardo and Marx. There was a general presupposition in this type of economic theorizing that the individuals in the economy were self-interested, but the main point of the theorizing—the long-run development of the capitalist system—was more about the interaction of different class-interests than it was about individual self-interest. It was the causes and consequences of the evolution of class interaction—three classes for Ricardo and two for Marx—that was the driving force behind most of the economic analysis. But more importantly here, it was often the case that the class-interest was directly opposed to the individual self-interest of the members of the class. For example, Ricardo (2015[1817]) recognized that it was in the individual self-interest of landlords to employ more capital and improve agricultural productivity, but while he recognized this point, it was not emphasized. Such agricultural innovations run contrary to Ricardo’s overall class-based story about the long run tendency of rents to rise and the rate of profit to fall, because such innovation would allow the society to produce more output on currently cultivated land. This would reduce the necessity to move to less fertile land, which in turn would reduce rents (or at least keep them from rising), a result that is directly opposed to the overall interest of the landlord class as well as directly opposed to Ricardo’s general argument about the falling rate of profit and the evolution of capitalist society toward the stationary state.

Similarly, in Marx (1909) there was also a tension between self- and class-interest although it concerned capitalists and capital rather than landlords and rent. In Marx the tension worked in a very different way. He argued that it was in the interest of an individual capitalist to increase the use of capital (dead labor) and reduce the use of labor (living labor). Such an increase in the productivity of labor will, in the short run, increase the profit of the individual capitalist. The problem is that once other capitalists do the same thing—and under Marx’s competitive assumptions they are forced to do the same thing—it would mean an increase in the overall capital to labor ratio (organic composition of capital). Since for Marx profit only comes from living labor, this will in turn reduce the rate of profit for the capitalist class. Given that for Marx the falling rate of profit is part of the historical mission of the capitalist class, this self-interested behavior, while obviously opposed to the class interest, plays an important role in Marx’s overall story about capitalist development.

Note the complexity and tension in these two stories. In both Ricardo and Marx certain individual self-interested actions were opposed to the general interest of the members of the class, and yet the consequences, for both the relevant class and the economy in general, were quite different. For Ricardo, the self-interest of landlords was relatively neglected in favor of his overall class-based story which focused on the increase in rents and the relative position of the landlord class, but with a result—the falling rate of profit—which was for Ricardo undesirable. On the other hand, for Marx the self-interested behavior of individual capitalists was emphasized as an important causal mechanism in his general class-based story about the development of capitalism and ultimate demise of the capitalist class. It seems rather ironic that the most pro-capitalist of the classical economists (Ricardo) would

down-play individual self-interest in favor of class-interest, while the most anti-capitalist classical economist (Marx) would have self-interest play key role in his analysis of the capitalist system.

As another example of the complex role that self-interest has played in economics, consider the neoclassical mainstream that dominated economics for most of the twentieth century. Such theorizing typically characterized individual decision-making in terms of *homo economicus*: the rational choice view that individuals have well-ordered preferences and act optimally on the basis of those preferences. This notion of rational behavior seems to depend on self-interest: choosing in such a way that the individual obtains their most preferred bundle and/or the bundle with the highest utility. In fact, as noted above, it has been a frequent criticism of rational-choice based economics that it always assumes self-interested behavior and thus denies the possibility that economic agents care about the well-being of others in any way. But rational choice theory does not require self-interested preferences; it only requires that preferences satisfy certain structural properties (completeness and transitivity) and does not restrict the content of the preferences: what the individual does or does not prefer. An individual can make rational choices to increase the well-being of others (altruism) as well as to decrease the well-being of others (malevolence). In fact economists frequently assume self-interested behavior, but it is an additional assumption and not a necessary condition for individual rational choice. So this is another case where self-interest plays a complex role and exhibits a number of tensions. On one hand, the idea that we all pursue our own self-interest seems to be the primary motivation for the economists' emphasis on rational choice, while on the other hand, choosing rationally, as economists define it, doesn't require choosing in a way that necessarily serves one's own self-interest. Daniel Hausman, Michael McPherson and Debra Satz summarize this tension very effectively:

To be self-interested is to have preferences directed toward one's own good, not simply to act on one's own preferences. What people who are not entirely self-interested judge to be best overall does not always coincide with what they judge to be best for themselves alone ... What distinguishes people who are self-interested from those who are altruistic or malevolent is *what* they prefer, and utility theory says nothing about the content of preferences. Utility theory leaves open the question of the extent to which individuals are self-interested (2017, p. 75).

Finally, the explosion of interest in behavioral economics during the last few decades also raises a number of questions about the meaning and role of self-interest in economics. Behavioral economics is primarily concerned with improving both the empirical track record and the psychological foundations of the theories that economists use to predict and explain individual choice behavior. Although behavioral economics has become an important field with many different concerns, the most important initial work was Daniel Kahneman and Amos Tversky's research on prospect theory (the literature is extensive but much of it is gathered in Kahneman and Tversky 2000). While behavioral economics has continued to develop various alternatives to traditional theory, its most enduring legacy has become the long list of well-known and persistent empirical anomalies to rational choice theory: loss aversion, preference reversals, framing, the status quo bias, hyperbolic discounting, and

such. As Richard Thaler put it: “The approach taken by most behavioral economists has been to focus on a few important ways in which humans diverge from *homo economicus*” (Thaler 2017, p. 1800). For the influential “heuristics and biases” program initiated by Kahneman and Tversky, these individual choice anomalies are considered *mistakes* in rationality. Individual decision-makers are generally assumed to have well-ordered preferences and desire to act optimally on those preferences—they have an “inner rational agent” (Infante et al. 2016)—but various psychological heuristics and biases often prevent them from actually achieving their rational goals.

While such behavioral anomalies and developing alternatives to rational choice theory remain the main thrust of much of the behavioral economics literature, my remarks will focus on a more normative aspect of the field—behavioral welfare economics—which re-examines various aspects of traditional welfare economics in light of the empirical anomalies identified by behavioral economics. The problem of trying to reconcile behavioral economics, which tends to undermine the long-standing economic presupposition that individuals act optimally on stable preferences, with normative welfare economics which has traditionally assumed the rational action of individual agents, is often called the reconciliation problem:

For at least the last three quarters of a century, both descriptive and normative economics have been based on assumptions about individual rationality ... however, there have been increasingly evident signs that economics might be changing direction, towards ... the *behavioural* approach. There has been an accumulation of work which tests rationality assumptions at the individual level ... and finds systemic ‘anomalies’ (that is, deviations from received theory). There has been growing interest in developing psychologically-based theories to explain these anomalies, and in applying these theories to various areas of economics. These developments pose severe problems for normative economics ... The problem of how to reconcile normative and behavioural economics – the *reconciliation problem* – is only just beginning to be recognized (McQuillin and Sugden 2012, pp. 553–554).

One popular approach to behavioral welfare economics has been the *libertarian paternalism* of Thaler and Cass Sunstein (Sunstein and Thaler 2003, Thaler and Sunstein 2009). Libertarian paternalism promotes a set of strategies and policies that nudge people away from making mistakes and into decision-making that will make them “better off, *as judged by themselves*” (Thaler and Sunstein 2009, p. 5). Since these strategies and policies generally operate by changing the context in which the choice takes place—the so-called choice architecture—in a heuristics and biases mitigating way, the individual is always free to make their original choices, they are just less likely to do so. In this way the program is both paternalistic (making people better off in ways they would not do on their own) and libertarian (the individual is still free to choose).

So how does libertarian paternalism raise issues and create tensions regarding the role of self-interest in economics? Most importantly Thaler and Sunstein insist that the successfully nudged agent will behave like *homo economicus*; those who behave rationally in their own self-interest are called “Econs” and those who have well-ordered preferences but fail for various heuristics and biases reasons, fail to act in their own rational self-interest are called “Humans” (Thaler and Sunstein 2009, p. 7). Libertarian paternalism offers various modifications to the choice environment that

leads Humans to act more like Econs without restricting the elements of the choice space or using traditional microeconomic policy tools like taxes and subsidies.

But this is a transformation of the traditional relationship between *homo economicus* and self-interest. Recall that economists have traditionally assumed that economic agents act rationally—have well-ordered preferences and act optimally on them—and also that economic agents are self-interested: their preferences reflect what they believe is best for them. But the many anomalies of behavioral economics have demonstrated in a convincing way that people often do not act in their own rational self-interest. This is of course the main critical aspect of behavioral economics as a science of individual behavior. Libertarian paternalism of course starts from the position that people make mistakes in rationality and self-interest, but instead of seeing this as reason to reject rational choice theory, it proposes a program to nudge such people into more rational behavior. This moves rational self-interested behavior from being what agents actually do—at least according to economics—to being a normative baseline for what people ought to do. As Mario Rizzo and Glenn Whitman explain:

The behavioral paternalist case hinges crucially on the idea that people deviate systematically from rational choice. Their policies are intended to push people toward more rational behavior. In other words, despite having rejected rationality as a model of how people *do* behave, the behavioral paternalist still accept rationality as a model for how people *ought* to behave (Rizzo and Whitman 2020, p. 16).

Unlike descriptive behavioral economics which shows how humans diverge from *homo economicus*, libertarian paternalism provides tools for how to get humans to behave more like *homo economicus*. This takes the tension-fraught relationship between rationality and self-interest discussed above and twists it in an another, entirely different, normative direction (see Hands 2020 for a more detailed discussion of libertarian paternalism and self-interest).

Of course these are just three of the many episodes in the history of economics which could be used to demonstrate the inadequacy of the conventional view about the relationship between self-interest and economics. These three examples were not selected because they were the main focus of any of the individual authors in this volume. They were selected because of their breadth (ranging over two hundred years of economics) and because of the complexity of the issues they raise (class interest and self-interest, self-interest and rational choice, and self-interest as a normative rather than a descriptive notion). The chapters in this volume demonstrate a variety of different ways that economics and self-interest have been connected in the history of economics and in particular how these relationships manifest themselves in the work of a number of specific economists.

The book is divided into three parts: the role of self-interest in the economics literature of the eighteenth and nineteenth centuries, its role in twentieth century economics, and a set of chapters raising more contemporary issues. All three sections are primarily historical although many concern philosophical issues—some in social and political philosophy, and some in epistemology and philosophy of science—and the majority focus on the work of a particular economist or school of economic

thought. For example, the chapters in the first section focus on the work of various pre-Smithian writers, Smith himself, Bentham and Mill, as well as Walras and Wicksteed. The second section on scholars like Otto Neurath and Karl Polanyi, economists such as Mises, Knight, and Hayek, as well as a chapter on questions about free will, political philosophy, and political science. The third section ranges even more broadly, from altruism, to an algorithm motivated by Smith’s *Theory of Moral Sentiments*, chapters on aspects of game theory and neuroeconomics, and a contemporary philosophical discussion of the relationship between economics, other disciplines, and social reality.

The volume offers the reader a vast amount of original, often philosophically-sensitive, research in the history of economic thought, as well as some more contemporary approaches to the self-interest/economics relationship. It is rich in depth and detail, with many authors—particularly in parts one and two—focusing on one significant economist or topic of importance. In total it represents a substantial challenge to the long-standing, and often polarizing, view of economics as simply the straightforward science of self-interest.

## References

- Hands, D. Wade. 2020. Libertarian paternalism: Taking econs seriously. *International Review of Economics* 67: 419–441.
- Hausman, Daniel M., Michael McPherson, and Debra Satz. 2017. *Economic analysis, moral philosophy, and public policy*, 3rd ed. Cambridge: Cambridge University Press.
- Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016. Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* 23: 1–25.
- Kahneman, Daniel, and Amos Tversky (eds.). 2000. *Choices, values, and frames*. Cambridge: Cambridge University Press.
- Marx, Karl. 1909. *Capital: A critique of political economy, Volume III, The process of capitalist production as a whole.*, ed. F. Engels. Chicago: Charles H. Kerr.
- McQuillin, Ben, and Robert Sugden. 2012. Reconciling the normative and behavioural economics: the problems to be solved. *Social Choice and Welfare* 38: 553–567.
- Ricardo, David. 2015[1817]. *On the principles of political economy, and taxation*, Cambridge: Cambridge University Press.
- Rizzo, Mario J., and Glen Whitman. 2020. *Escaping paternalism: Rationality, behavioral economics, and public policy*. Cambridge: Cambridge University Press.
- Smith, Adam. 1776. *An inquiry into the nature and causes of the wealth of nations*. London: W. Strahan and T. Cadell.
- Sunstein, Cass R., and Richard H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review* 70: 1159–1202.
- Thaler, Richard H. 2017. Behavioral economics. *Journal of Political Economy* 125: 1799–1805.
- Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving decisions about health, wealth and happiness*. London: Penguin.

# **Self-Interest in Eighteenth and Nineteenth Century Economics**

# Self-interest and French ‘Philosophie économique’ 1695–1830



Gilbert Faccarello and Philippe Steiner

**Abstract** In this chapter, we focus on the French economists that, all along the eighteenth century, formed what we call “philosophie économique”. In Sect. 2, this current of thought is precisely defined: a review of the troops shows how its members (Boisguilbert, Quesnay, Turgot, Condorcet and Say) refer to a new view on the nature and role of self-interest, and why, in their opinion, self-interest is supposed to reach positive results, both at the individual and collective levels. Section 3 deals with how, in this approach, the Legislator is supposed to act in this new environment, trying to use self-interest as a means of government, and how it is supposed to make decisions. Section 4 concludes, stressing the fact that, parallel to the recognition of the positive role of self-interest, more and more critical voices arose to stress at the same time the limits of this approach and the essential role played by other important elements—religion, morals, altruism—neglected by “philosophie économique”.

**Keywords** Boisguilbert · Competition · Condorcet · Quesnay · Philosophie économique · Say · Self-interest · Turgot

## 1 Introduction

From the second half of the reign of Louis XIV to the French Revolution, economic matters progressively became a central topic in politics in Europe, first in England, and then in France. In France in particular, the number of publications of books and pamphlets in this field increased dramatically during the second half of the eighteenth century (Théré 1998). Even works like the celebrated *De l’Esprit des Lois* (1748), by Charles de Montesquieu, or the *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* (1751–1772), edited by Denis Diderot and Jean Le Rond d’Alembert, included developments on economic questions. Authors began aware to

---

G. Faccarello  
Université Panthéon-Assas, Paris, France

P. Steiner (✉)  
Sorbonne Université Maison de la Recherche, 28, Rue Serpente, 75006 Paris, France  
e-mail: [philippe.steiner@paris-sorbonne.fr](mailto:philippe.steiner@paris-sorbonne.fr)

discuss the new field of knowledge and the sentiment of novelty lasted for a long time, at least until the first decades of the nineteenth century and the works of Jean-Baptiste Say, when the new intellectual discipline seemed to stabilise. While all kinds of economic subjects were debated (taxes, public expenditure, money, banking, prices, foreign trade, etc.), the main subject of controversy, central in almost all discussions, was *laissez-faire* and the alleged positive role of self-interest. To assert the role of self-interest was a genuine intellectual revolution that economic ideas provoked in morals and politics, two fields in which it had almost always been held as self-evident that people should fight their selfish sentiments and governments should put some strong limits to the interested behaviour of the citizens.

However, as it is easy to guess, over more than a century and a half the developments in economic thought were not homogenous. Among the wealth of the literature of the time—and leaving aside the writings of authors who fought against the modernity advocated by free trade in the name of some tradition, or who proposed another kind of modernity—it is possible to distinguish two main currents of thought (Demals and Faccarello 2016): “commerce politique” and “philosophie économique”. The first consists in a French adaptation of the English “science of trade” and is illustrated by such authors as Jean-François Melon, Nicolas Dutot, Montesquieu, or the members of the circle of J. C. M. Vincent de Gournay (Charles, Lefebvre and Théré 2011)—the main figure of the group being François Véron de Forbonnais. The second includes those who fought more decisively in favour of the “liberté du commerce”, from Pierre Le Pesant de Boisguilbert to François Quesnay, Anne-Robert-Jacques Turgot, Marie Jean Antoine Nicolas Caritat de Condorcet and, at the end of the period, Jean-Baptiste Say. Both currents of thought proposed new political philosophies centred on new economic policies, that is, more or less radical ways to change French politics and to reform the State.

In this chapter, we focus on the second, more radical group, that is, “philosophie économique”. In Sect. 2, this current of thought is precisely defined: a review of the troops shows how its members refer to a new view on the nature and role of self-interest, and why, in their opinion, self-interest is supposed to deliver positive results, at both the individual and collective levels. Section 3 deals with how, in this approach, the Legislator is supposed to act in this new environment, trying to use self-interest as a means of government, and how it is supposed to make decisions. Section 4 concludes, stressing the fact that, parallel to the recognition of the positive role of self-interest, more and more critical voices arose to stress the limits of this approach and the essential role played by other important elements—religion, morals, altruism—neglected by “philosophie économique”.



## 2 The Logic of Self-interest from Boisguilbert to Say: Self-interest, Markets and the Competitive Process

By “philosophie économique”—a phrase coined by the Physiocrat Nicolas Baudeau (Baudeau 1771)—we understand a series of authors who, from the end of the seventeenth century to the 1830s, developed a new way of understanding the functioning of the economy, and hence to rule a society in which economic affairs are more and more important (Faccarello and Steiner 2008a). This new approach is organised around three elements. Firstly, these authors uphold a self-interested conception of human behaviour on a pragmatic level as daily action directed to monetary profit, as well as, on a purely intellectual level, various forms of reasoning stressing the role of *utility*—a utility which was supposed to condition and legitimate action. This stress on utility is either different from Bentham’s better known but subsequent utilitarianism—Quesnay, for example, connected a theory of natural law to utility-motivated actions, and Turgot insisted on the rights of men while developing a subjective theory of value—or close to it as it is the case with Say. Secondly, adopting the sensationalism developed by John Locke and more specifically by Étienne Bonnot de Condillac they relied on a theory of knowledge that explained the way in which individuals grasp the world intellectually. Thirdly, there is a link to the Legislator. In France, the process of centralisation that lies at the core of the Bourbon monarchy and its administration, together with the rationalism inherited from Cartesianism and sustained in the eighteenth century through the lasting influence of the philosophy of Nicolas Malebranche, gave a salient role to the state even if some authors placed greater emphasis upon the representation of interests, or the implication of enlightened citizens in a decentralised political organisation.

“Philosophie économique” seems to deal only with the economy: however, it was dealing with politics as well. This was particularly clear in the second part of the eighteenth century when the Physiocratic movement proposed political views grounded on self-interest in contradistinction to Montesquieu’s positive evaluation of “honour” in a monarchy and Rousseau’s emphasis on the role of “virtue” (Charles and Steiner 1999). Self-interest was thus an intellectual tool for building a bridge between the economy and politics.

Who were the most important proponents of “philosophie économique”? Boisguilbert (1646–1714) is certainly the first to be mentioned for his contribution to the foundations of political economy at the turn of the seventeenth and the eighteenth centuries. Then comes François Quesnay (1694–1774) with the group known as the Physiocrats, together with Turgot (1727–1781) and Condorcet (1743–1794), both notably influenced by Quesnay as regards the exclusive productivity of agriculture, but firmly standing on their own with the development of more personal views of political economy. They were active from the middle of the 1750s until the French Revolution. During this period of time, the utilitarian morals and political philosophy of authors like Claude Helvétius (1715–1771) and Paul-Henri Thiry d’Holbach (1723–1789) became influential and strengthened the idea that self-interest was the leading spring of any human action. This radical moral view was rejected by Turgot,

Condorcet and the Physiocrats, who did not accept the demise of deontological moral views. However, at the beginning of the nineteenth century, Say (1767–1832), the last significant member of this current of thought, advocated utilitarianism in its Benthamite form. Finally, “philosophie économique” ended in the 1830’s when the role of non-utilitarian dimensions in economy and society—including the religious dimension—challenged the self-interested views developed by “philosophie économique” (Faccarello and Steiner 2008b).

## ***2.1 The Emergence and Development of the Logic of Self-interest***

Boisguilbert, at the end of the seventeenth century, was the first to mark out this position on a Christian basis, elaborating on some Jansenist ideas stated, in particular, by the theologian Pierre Nicole. He argued that if one was to uncover an order within economic activity, it was enough to consider the motivations of agents, which are nothing but the translation into economic life of the fundamental selfish conduct of men dictated by the Original Sin and the Fall of Man: “each thinks of attaining his own personal interest to the highest degree and with the greatest possible ease,” he writes in 1705 in his first *Factum de la France* (Boisguilbert 1966, p. 749). What is particularly striking here is that Boisguilbert, starting from Jansenist religious thought (Faccarello 1986), explains how a basic passion like cupidity could be neutralised in markets with the assistance of a specific social mechanism: free competition. By confronting each individual’s cupidity with the cupidity of all other people, competition eliminates the socially harmful effects of selfishness and makes an orderly society possible, a “harmony”, as if each individual were motivated by charity—when in fact he is not. Boisguilbert’s approach was taken up by the other contributors to “philosophie économique” after some decades: Quesnay, Turgot and Say developed it in various but complementary ways—discarding however its theological basis and replacing it with forms of sensationist philosophy.

In his first economic writings, Quesnay compared the then current situation of restricted grain trade with a free market in order to explain the positive outcomes of free competition. In the situation that prevailed at that time, prices were strongly correlated to the volume produced since they were following a King-Davenant law, with very low/high prices following good/bad harvests (Steiner 1994). This situation was unsatisfactory because when the customers enjoyed low prices, the farmers were experiencing losses; whereas when farmers were making a nice profit, the buyers were experiencing high prices. In such a situation, interests did not converge and all economic actors were suffering. Had economic policy been based on free trade, then, with the same average prices over a period of five years, interests would have converged since the King-Davenant effect would have disappeared: low prices would have been associated with good harvests and profits for the farmers, and high prices with bad harvests and losses for the farmers. This approach was taken up by Turgot:

in his 1770 *Lettres au contrôleur général*, he developed the case for free competition as the best means to cover the needs of the population, instead of the usual policy of market restriction and public storage (Faccarello 1998). Connecting Boisguilbert's "price of proportion" with Quesnay analysis of the grain trade, he claimed that the customer is mainly interested in the stability of prices since wages on the labour market are then "in proportion" with the price of grain. From the point of view of farmers, free trade and the stability of prices are important as well since the profit they receive allow them to invest and increase their production. If local shortages were to happen, and as "the needs of the customer increase the price of a good as it is less abundant, it follows necessarily that there is a great interest in bringing grain from places where it is abundant to places where it is scarce" (Turgot 1913–23, III, p. 322).

Later on, Say took a broader view, going beyond the grain trade in order to promote self-interest *and* competition as the ultimate political tool for achieving the best economic state of well-being in a society: "Self-interest is always the best judge of the amount of revenue that one may expect from production; and even if self-interest may be mistaken, it remains the least dangerous judge, and the one that costs the least. But self-interest does not convey any information when self-interests do not compensate each other" (Say 1825, I, pp. 213–214). As for Turgot, self-interest was thus fully recognised as the driving force in a market economy and, according to Say, the real issue was to acknowledge the role of competition as the necessary policy to turn a self-interested behaviour into a powerful force to realise the general interest (Steiner 2006). The opposite situation is given by the spread of "vanity". "Interest, so vilified, does not bring us to making so great mistake [religious punishments, wars, etc.] as we do following vanity, which a contemporary philosopher calls anti-social vice. The interest of a man is not always, and even not often, opposite to the interest of another man. Every useful worker is beneficial to himself and to those for whom the work is done; but vanity is necessarily in opposition with the vanity of other men, since one cannot dominate without the other be subjugated" (Say 1817, pp. 89–90). Spreading political economy among public opinion and fighting vanity within the government is thus an important political task since knowledge of this science is the best way to know one's own true interest and to stick to it for the benefit of all.

Beyond these considerations on the competitive process and the positive role of self-interest in the grain trade and, in general, in society, a free foreign trade policy was a key issue in the development of "philosophie économique". While free trade was essential to the realisation of a system of competitive prices that ensure the compatibility of opposed interests in markets, these authors, however, went further: they all linked free trade in domestic markets to the freedom of foreign trade. Free trade at home, they argue, can only stabilise the price of corn and create a condition of wealth based on a harmonious system of relative prices if it is supported by a free foreign trade, whatever the state of the balance of trade.

This new political view of foreign trade, initiated by Boisguilbert, is important for another reason: it provided a solution to the problem caused by the diverging material interests of different countries. These interests could be peaceably harmonised if the merchants were able to trade freely in international markets, pursuing their own

private interests. Governments must thus understand the main conditions for the realisation of a state of domestic harmony and plenty: comprehensive free trade, internal and external. In the second half of the eighteenth century, “philosophie économique” developed especially this line of thought in the works that aimed to disseminate the Physiocratic doctrine. Pierre-Paul Le Mercier de la Rivière and Guillaume-François Le Trosne, for example, explicitly presented the idea as a political alternative to the policy of a “balance of powers in Europe”—a policy they considered to be a fertile source of disagreement and warfare between States. “The principle of fraternity of nations is not ... only dictated by justice, but it also accords with the interest of each nation, independently from the behaviour of the others. It should not simply be regarded as a beautiful moral idea, a worthy conception to be taught in schools of philosophy, but also as a practical maxim of government from which we can only detach ourselves to our own detriment” (Le Trosne 1788, pp. 413–414).

This idea caught on. Say stated it, for example, in his *Cours complet d'économie politique pratique*, developing the international consequences of the so-called “law of markets”: “What is true about one individual in relation to another is also true for a nation in relation to foreign countries; each one is interested in the prosperity of all the others, as one can only sell to nations which are in a position to buy; and a nation ... can only buy with that which it produces. This more exact conception of the nature of things now holds sway over the enlightened views of a few nations, and should progressively change the policy of the world; because as men become more enlightened, they listen to the suggestions of their enlightened interest, which for them are more reliable than the dreams of philanthropy.” (Say 1828–9, VI, pp. 317–318).

## 2.2 *Removing Some Obstacles to the Positive Role of Self-interest*

Self-interest, in a competitive environment, is thus supposed to be always beneficial and lead to the general welfare of society. Whenever economic private interests are opposing each other, free competition forces them to converge to an optimal individual and collective situation. This is what economic theory states or should state: “it is always the best with which one must be concerned in theory. To neglect this research, on the pretext that this better is impracticable in current circumstances, is to wish to resolve two questions at the same time: it is to give up the advantage of raising questions in the simplicity that alone can allow them to be demonstrated” (Turgot 1913–23, II, p. 294). But “philosophie économique” also considers situations where, in practice, some prejudices or vested interests can raise obstacles to the competitive process.

An important problem that could prevent the beneficial effects of the interplay of economic private interests in competitive conditions deals with the mentalities of the people—what is now called “moral economy” (Thompson 1971). On two points especially, these attitudes were deeply held within the population and among rulers

for centuries: namely, the regulation of the rate of interest, and the traditional negative image of the merchant. For centuries, usury had been a subject of heavy discussions in religion and moral philosophy, and the Legislator, since Charlemagne, provided support to the Church in forbidding the lending at interest: the lender was a usurer whose self-interest was supposed to lead to high interest rates, damage the borrowers and ruin their families. As for the merchants, especially in the corn trade, they were supposed to be bare speculators, monopolising corn and selling it at very high prices when people were starving. On these two very sensitive issues, Quesnay and Turgot for instance showed that people and the legislator should not fear free trade and the self-interested actions of the agents. On the first point, from Boisguilbert to Say—implicitly for Boisguilbert and openly for the successive members of “*philosophie économique*”—it was said that the determination of the rate of interest must be free: it was a necessity for free trade and the development of economic activities and national wealth. Debunking the main arguments of the Scholastics in favour of the prohibition of the lending at interest, Turgot showed that a positive rate of interest cannot harm borrowers or lenders if it is freely debated and determined in competitive conditions: this is the first economic theory of the rate of interest, based on Turgot’s theory of subjective value (Faccarello 2016a).

On the second point, Quesnay and Turgot insisted that, whenever competitive conditions prevail, merchants cannot manipulate markets and are, on the contrary, an essential cog in the general machine of prosperity: buying where corn is abundant and cheap, bringing it to markets where it is rare and expensive, thus regulating and stabilising prices. To benefit from free competition in markets, it is thus necessary that merchants be rehabilitated in the eyes of public opinion, and protected by the public force. Otherwise, whenever the profession of merchant is scorned, those who practice the trade will see regularly their wealth subject to confiscation by crowds during upheavals, so rich and honest persons will never enter this profession. Thus the grain market will witness regular crises, disorganising the rest of the economy. This is one of the reasons why, because it shared the prejudices of the crowd or for fear of unrests, the Legislator tended to regulate the corn trade. “Do we believe that, by hindering trade through degrading difficulties, by intimidating warehousing, by announcing that one regards grain to be less sacred as property than any other asset ... we will ensure that more of it will be warehoused? If there were people who still entered this trade, doubtless they would count these new risks and their shame among their costs, and they would make the consumers pay for them. There will not, however, be any of them, for to carry out this trade in such a way as to fill the needs of a suffering people, there must be huge advances, large capital, and rich and accredited merchants; now, the latter are not made to record to a registry of police; they do not put their fortune at the mercy of a judge, or even of the government” (Turgot 1913–23, III, p. 323). When he was in power, one of the first things that Turgot did was to order the public force to protect the merchants.

Slavery is another important issue that illustrates the role of self-interest in the presence of violence. Slavery begun to raise moral concerns in the 1750s following

Montesquieu's critiques in *L'Esprit des lois* and some entries published in *Encyclopédie*. In 1771, two papers written by Pierre-Samuel Du Pont in the Physiocratic *Ephémérides du citoyen* connected this moral issue with economic calculation (Oudin-Bastide and Steiner 2019, Chap. 1). The main theme of his papers was to show, through an economic calculation similar to Quesnay's calculations of the cost of production of corn, that the cost of slave labour was higher than the cost of free labour. Accordingly, slavery that was obviously detrimental to the interest of the enslaved persons was also detrimental to the interests of the colonists. Some years later, in his 1781 pamphlet *Considérations sur l'esclavage des nègres*, Condorcet did not endorse Du Pont's calculation, considering that slave labour was cheaper than free labour; but this significant difference did not prevent him to act in favour of the abolition of slavery because, he said, the interests of the nation and of the slaves were more important than the interest of the colonists alone who, in the long run, would as well benefit from the economic development that would necessarily follow the abolition of slavery and the increased productivity of free labourers in the French West Indies.

Finally, another problem could sometimes prevent the smooth working of free markets: the existence of uncertainty, for example in international—and more specifically in maritime—trade, and in the related questions of insurance. Condorcet thought that this basic problem was in fact concerning all kinds of activities: in all fields of life where decisions are to be taken, he stressed, people almost always have to face uncertainty. He thus conceived of any economic activity as an uncertain and risky undertaking—"undertakings in which men expose themselves to losses in view of a profit" (Condorcet 1994, p. 396). Himself a mathematician, he thought that the development of the calculus of probabilities could help individuals in their choices, and he started to use probability theory to describe the entrepreneurs' decisions to invest (Rieucan 1998; Faccarello 2016b). Self-interest was thus supposed to find in this new branch of mathematics a powerful tool for decision support and calculation.

### 3 Self-interest and the Legislator

All these developments were supposed to show that a self-interested led economy performs better than a "war economy" or a "moral economy" advocated by those who opposed "philosophie économique". But, our authors stressed, this self-interested economy includes a state—an important point that must be addressed now. In the first place, "philosophie économique" considers how the state produces the necessary underpinnings for such a self-interested economy. In the second place, it develops the idea that the government can use the self-interested behaviour of the economic actors in order to reach its goal. Finally, beyond these practical aspects, it also deals with some normative issues on how the Legislator should make its decisions in order to avoid the traps of its own interests.

### 3.1 *Producing a Self-interested Economy*

The basic rule proposed by Boisguilbert and taken up by the authors after him, is that the Legislator should ensure a smooth functioning of competitive markets—the condition for the convergence of private and public interests. In this perspective, its main task is to provide a sound legal and political environment for free trade, that is, to protect economic agents from any violence, domestic or foreign—in other words, the Legislator must provide the services of police, justice and defence. “We said: provided that nature is left alone, in other words, that nature is given its freedom, and that one only ever intervenes in order to provide protection and prevent violence” (Boisguilbert 1966, p. 892). However, Boisguilbert stressed, government should not only prevent itself from intervening in markets with any kind of regulations—all the existing regulatory decrees or laws should be repealed—but also from intervening in markets as an economic agent itself, for example buying and selling corn with the aim of stabilising its price (except when a war is impeding a free foreign trade). The members of the government do not have the relevant knowledge of the necessities and constraints of economic life and their interventions inevitably distort the functioning of markets. This means that the government should also be wary of its own benevolent sentiments—not to speak of the pressure of vested interests. It often believes that, in regulating economic life, and especially in putting strong constraints on merchants and the corn trade, it has done the right thing to alleviate the poor. But in fact the opposite is true. Men in power, Boisguilbert stated, should fight their own self-love and recognise they were wrong in acting in such a way. “Self-love, especially when placed in an eminent position, is reluctant to understand that it has been grossly mistaken for a very long time ... What is most troublesome is that the opposite party claim to have love or pity for the poor on their side and the most just measures to avoid the terrible effects of violent infertility ... What a strange destiny it is for the monarchy, in order to be rich and happy, to have only to defend itself from the zeal and good intentions of those who manage it!” (Boisguilbert 1966, pp. 335–336). These statements were the basis of the *laissez-faire* policy that was endorsed by all those participating in “*philosophie économique*”, from Quesnay to Say. Self-interest must be the leading driving force and the government should significantly reduce his grasp on the economy and society. This self-interested view of the relation between the government and society could have far-reaching consequences. At the end of our period, after the episodes of the Revolution and the Empire, in one of his 1819 public lectures at Athénée, Say could go so far and, quoting the Italian motto “*Il mondo va da se*” (the world goes by itself), insisted that a country could perfectly live without a government—police and justice being provided by the citizens themselves or by private bodies (Steiner 2006; Faccarello 2010).

But the existence of a state as the producer of services of police, justice and defence is only part of a wider and necessary set of functions whenever the logic of self-interest in markets proves to be inefficient—this is the well-known question of the public goods, merit goods and externalities, that, from Turgot onwards, was



discussed in “philosophie économique” (Faccarello 2006). Thus, producing a self-interested economy did not mean that, in some circumstances, the Legislator should not interfere with the self-interests of the citizens. Let us take some examples.

In “philosophie économique”, a key concept was of course that of private property rights. It is true that Boisguilbert, at the end of the seventeenth century, did not pay special attention to such rights. Later, Turgot, by contrast, well aware of the developments of natural right theories and of the debates in the field, stressed that the (private) property right, while naturally originating in the necessity of withdrawing products out of the original joint-possession in order for them to be useful to individuals and to increase productivity (lands would not have been cultivated in its absence), is only a social convention aimed at the general utility of the community. It is thus justified by the general wealth it generates, that is, the positive benefits for society. For this very reason, Turgot stressed, it is by no means an absolute right and limits should be set on it: the Legislator can and must intervene every time this justification no longer exists or fades away, even at the price of hurting the self-interest of the citizens. The legal definition of the property right as the right “to use and abuse” was thus thought as inadequate. Take for example the case of a landowner. “Natural right seems to allow everybody the entitlement to dispose of his land at his discretion. This is a consequence of the property right ... but ... one cannot doubt that the legislator is entitled to regulate this usage with the view of the general utility of society; to prevent for example a man from substituting a sterile magnificence to a useful fertility, and a piece of land aimed at sustaining a group of workers from being only devoted to the promenade of some leisured men” (Turgot 1913–23, I, p. 439).

This approach, which allows the State, under specific conditions, to interfere with private interests, also finds an expression in what is now called “merit goods”. Here the Legislator clearly dictates, in the name of the public interest, what citizens should do or avoid doing. This is the reason why, for example, Turgot and Condorcet wanted to oblige people to send young children to school: instruction was seen by them as a fundamental good for the education of the citizens. Say endorsed that view as well and, in the first edition of his *Traité d'économie politique*, suggested a general law for this kind of state interventions curbing the way people act: “In these cases, government must decide nothing but what citizens would have done had they the knowledge that the government has” (Say 1803, I, p. 251).

This raises the more general question of the externalities, positive or negative, at that time a widely debated theme, the symbol of which, during the eighteenth century, was the problem of the existence of marshes and the negative effects they generated on their surroundings—spreading malaria among the population and preventing pieces of land to be cultivated. Here also the public interest must oppose the self-interest of the landowners, and, whenever the owners of marshes did not want to drain them because they prefer to use them for watermills or for fishing and hunting, the Legislator is entitled to force them to sell him the land in order for the State to drain it, make it cultivable and finally resell it to private cultivators. Turgot and Condorcet were unambiguous on this point.



### 3.2 *Where the Legislator Uses Individual Self-interest to Reach Its Goals*

“(Economic government” (“gouvernement œconomique”), a core concept of the Physiocratic theory and one of Quesnay’s linguistic innovation, makes a link between government and the economy (Steiner 2014). It refers to the management of real wealth, that is, the net product generated by the agricultural sector of the economy, and thus the general interest: “The status of population and employment of men are thus the principal objects of ‘œconomic government’ in any state, because it is from the work and industry of men that the fertility of the land, the monetary value of its production, and the good use of monetary wealth result. Here are the four sources of abundance; they each concur in their mutual growth, but they cannot do so without the general administration of men, goods, productions” (Quesnay 2005, p. 259). Administration is thus also considered in relation to self-interest. The link between administration, “œconomic government”, and a policy grounded on the self-interest of citizens is particularly strong where the reform of the taxation system is at stake.

Quesnay’s theory of wealth follows from the question of who should pay taxes, and how taxes should be paid. “A long time ago, it was said that taxes in themselves do not ruin the State, but instead the way that they were imposed. This has been understood, but nobody has sought to prove or elaborate this truth ... As a matter of fact, it would be necessary to know in detail the sources of the annual wealth of the state in order to set a proportional tax upon the wealth of subjects” (Quesnay 2005, p. 226). So while Quesnay’s position on wealth and taxation is clearly in line with the opinions of the other authors of “philosophie économique”, the administrative dimension of the issue is of significance to him, as he stated explicitly when commenting on Jean-Baptiste Naveau’s *Financier citoyen*: “It is ... the immense sum paid to the lower ranks of the administration, the injustices and trouble in commerce, that prevent the beneficial progress that the author imagines ... This multitude of clerks busy with levying taxes upon goods is, in itself, a loss of men to the State: all these men paid by the nation do not produce by their work any wealth for the State, thus this expenditure and these men are nothing but a sheer loss to the kingdom” (Quesnay 2005, p. 230). The connection between administration and “œconomic government” makes it necessary for Quesnay and the Physiocrats to seek a possible solution to the issue, and to propose a way in which the system of taxation should be properly administered without the trouble and the huge costs associated with the prevailing fiscal administration. The farmers’ immediate self-interest formed the basis of their fiscal policy.

It is well-known that the Physiocratic solution to the problem of taxation is to levy a single tax directly upon the net product—commercial and financial revenues being thus exempted from taxes. This straightforward solution clarifies matters, since objections to inquisitive administrative practices become irrelevant. However, Quesnay and Victor Riqueti de Mirabeau stressed, to tax the net product raised a difficult question: how is it possible to know, not the value produced on a piece of land but the net product obtained on that land, without relying on local administration and

its politically unacceptable behaviour? Mirabeau brought an answer to this problem. In his opinion, there was no need to seek access to the “secrets of the households” to establish the exact amount of the net product: the existing leases provided the needed information. “In a word, a lease is the only reliable basis for the government when assessing the net product, that is, the revenue of the country. It is only from this revenue that the government can take its share; the total of the leases is the total amount of the revenues of the nation, and it is on this sum that the share of the State can be levied. ...; this is the only way to achieve an enlightened agreement between the king and the nation, the necessary basis for good administration in the domain of finance” (Mirabeau 1761, p. 248)

The market agreement between the landowner and the farmer is the solution, since the agreement states explicitly the amount of rent to be paid by the farmer to the landowner. Is this solution a decisive step in ending the obstruction of administration? Was not Mirabeau naïve in believing that the actual leases will be accessible to fiscal administration? He made clear to the reader that he was aware that leases were already the basis of the calculation when the ‘vingtièmes’ (5% tax) and the ‘dixièmes’ (10% tax) were levied; and that the real leases were hidden, and only false data given to the fiscal administration. Nevertheless, Mirabeau claims, the case of the single tax on the net product is different: assessed through the leases, it offers a solution because it is in the direct interest of farmers implementing the “grande culture”, or the capitalist form of husbandry—furthermore, according to the Physiocratic theory of the incidence of taxation, it is also in the direct interest of the landowners. These capitalistic farmers would not jeopardise their wealth by using counterfeit leases; they would have no interest in cheating since the taxes would be paid by the landowner, and not by themselves: “[farmers] would not conceal the agreement, notably because they have no interest in the diminution of a tax which does not concern them” (Mirabeau 1761, p. 252). Nor would the farmers be ready to cheat and then share with the landlord a portion of the taxes due to the government. Such a strategy of tax avoidance would imply a complex system in which a counterfeit lease would have to be written to diminish the rent paid to landowners and then their taxes, and a part of that sum—a third in Mirabeau’s example—would be left in the hands of the farmer. In exchange for that, Mirabeau supposed that the concealed part of the lease would be paid in advance to the landowner, so that the latter would be sure to get his full share. Such a system would cost farmers too much capital. Consequently, Mirabeau claimed, the leases used would be the real ones, thus providing an accurate basis for the single tax.

The developments proposed by Turgot, Condorcet and André Morellet (Faccarello 2006) on the question of the tax assessment and collection are a bit different. In particular, they also tried to find a solution to the possible opposition between the public and the private interests, and they relied on the dynamics of self-interest. However, it is important to notice that in this case, unlike that described by Mirabeau, the relevant interests are not those expressed in markets: instead, they oppose each other through protests and dealings with the administration. As a matter of fact, two different systems were used in France. One, the “impôt de quotité”, required the payment of a given percentage of the revenue: the landowner should pay, every year,

this percentage out of the “produit net” of his land. A second system, the “impôt de répartition”, consisted first in determining the total amount of money needed for public expenditure, and secondly in distributing this amount across taxpayers according to their income—in the present case the “produit net” of land. Of course, the two systems were by no means socially and politically equivalent. But two aspects of the “impôt de répartition” were particularly interesting to our authors, and both were stressed by Turgot in 1777 in a letter to Benjamin Franklin, and later developed by Morellet (1790) during the French Revolution. In the first place, this kind of tax includes, so to speak, a built-in mechanism preventing people from infringing the law: whenever a taxpayer is tempted to hide part of his income, thus increasing the share of taxation supported by his fellow citizens, the interest of the latter is to protest and to make him pay his appropriate tax amount in order for them not to pay for him—a reaction that does not exist with the “impôt de quotité”. In the second place, thanks to this mechanism, a less extensive knowledge of the income and wealth of the taxpayers is required. “With the first system [“impôt de quotité”], the State ... finds itself alone against all the citizens who are all equally interested in disguising their income. With the second system [“impôt de répartition”], the citizen who would like to hide his income finds himself alone against all others who are interested in discovering it in order not to be burdened with his share ... With the first system, the government must know the absolute value of the income of each landowner, because it has to receive a given percentage of this income. With the second system, one is seeking less to know the absolute value of each estate than its relative, or comparative, value” (Turgot 1913–23, V, pp. 519–20).

### ***3.3 The Legislator, Public Opinion and the Process of Decision Making***

The problem of public decision-making had always been at the centre of the debates during the period under examination, and “philosophie économique” is no exception. Who is making public decision? How, and on what basis? How should the Legislator overcome prejudices and resist the pressure of self-interests—be they that of intermediary bodies, influential private persons, or its own interest? The spread of political economy and the search for political arrangements that could produce policies favourable to the general interest were the main proposals coming from the members of “philosophie économique”.

At the end of the seventeenth century, Boisguilbert again put to the fore a very old complaint: that the King and the government are surrounded by advisers, courtiers and sycophants, who, belonging to the leisure class, have no knowledge of economic life and thus tend to confirm the Legislator’s prejudices in this field. Moreover, they also have a strong tendency to suggest measures favouring their own private interests. It is thus easy to understand that, in these conditions, and to borrow a judgment of a Jansenist author esteemed by Boisguilbert, the life of the rulers “is no more

than a dream in which they see nothing but false objects and misleading phantoms” (Nicole 1670, p. 23). How to get out of this difficulty? Boisguilbert suggested two main measures. First of all, the government should systematically be in search of information and competence coming from the various kind of agents themselves: this, he stressed, was the role of the General Estates of the Realm, when the three estates could inform the Monarchy on the real condition of the country. Unfortunately, the General Estates are no more convened. This is the reason why a second best is to listen to what the Parliaments—although belonging to the privileged class—have to say (at that time Parliaments were provincial courts of justice, the Paris Parliament having in addition the role of registering Royal laws and decrees and the possibility to admonish the government). Second measure to be taken: the royal administration must be made of persons who understand economic life and are able to inform the government on the real economic conditions of the provinces and the working of markets—in other words, they should be chosen among the Third Estate, and more precisely among those who had been personally successful in their own private professions (Boisguilbert 1966, p. 297).

Quesnay and the physiocrats were as well deeply interested in the public decision-making. In this respect, the physiocrats endorsed Quesnay’s views that the spread of political economy would have been a significant step to get an enlightened public opinion, that is, an opinion capable of knowing what a good economic government is and where its true self-interest lies. This would be a powerful means to the convergence of the self-interests of the different social classes (Charles and Steiner 1999, pp. 85–86). Hence, it is no surprise to read, in the second maxim of Quesnay’s 1767 *Maximes générales du gouvernement économique d’un royaume agricole*: “The nation must be well educated in the general laws of the natural order that form evidently the most perfect government ... Those who aim to belong to the administration must learn the natural order the most useful to men in society” (Quesnay 2005, p. 566). Nevertheless, the Physiocratic school went beyond this general statement about the spread of the “new science of political economy” with two different political proposals. The first one came with the theory of legal despotism proposed by Quesnay and then developed by Le Mercier de la Rivière (1767): it rested on the formation of a body of experts in full command of political economy, so that they would check whether a public decision is congruent with the statements of the science. This boiled down to a form of constitutional verification of the relevance of public decisions. In Le Trosne’s words: “Nobody but an authority enlightened as to its true and well calculated interest can intervene in private interests” (Le Trosne 1788, I, p. 129). The second proposal was the creation of assemblies in which the landowners would take major decisions, notably those related to the amount of taxes. This was particularly developed by Mirabeau, and then by other members of the group, such as Du Pont. The landowners had to play a prominent role because the Physiocrats considered that their interest was the same as the true general interest. “Generally speaking, interest is and must be the true and long lasting spring of action of men; and the more their interest is close, present and noticeable, the more their action will be sure and steady. Public interest, that is, a good order, is infinitely closer to the class of landowners than to any other class” (Mirabeau 1769, p. 183). This important

idea is reassessed in the eulogy that Le Trosne would have liked to hear from the landowners' mouth: "Our interests are the same as Yours [the King's interests]; and this common interest consists in getting the largest output from the land ... It is time to overhaul the present management system, which makes You poorer and which makes our efforts fruitless" (Le Trosne 1788, I, pp. 146–147).

During the 1760s and 1770s, Turgot also took seriously the question of the decision-making processes that the Legislator should follow to implement the right economic policy, and, in general, in any decision-making concerned with police and justice, in order to avoid choices dictated by prejudices or vested interests (including its own) and act in favour of the general interest of the country. As regards an economic policy in the public interest, Turgot thought that the right guide to action was political economy, and he himself developed economic theory in an important and decisive way. More specifically, in a correspondence with Condorcet he discussed at length the question of how a decision should be taken through a voting procedure. Even if, when he was in charge of public responsibilities, he sometimes felt that, to rule and reform a country, it was much easier to have full legislative and executive powers in order to avoid wasting time in endless discussions with prejudiced men and vested interests, he planned an important reform aiming at transferring most of the powers of the King to a series of elected assemblies—municipal, provincial, national. It is true that, according to the Physiocratic scheme that Turgot accepted for its most part, voters would only have been the landowners. But this notwithstanding, this explains why he was concerned about the best way for an assembly to vote and make decisions that would have been in accordance with the public interest. He did not have time to implement his reform and to develop his ideas, but Condorcet, after him, developed original views on this important theme, clarifying Rousseau's concept of "general will". In his *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (Condorcet 1785), Condorcet tried to find out a set of voting rules that would have brought collective decisions which could be at the same time democratic and "true", that is, in the case of economic decisions, compliant with what the theory states, avoiding the traps of all kinds of private interests. One well-known result was called later the Condorcet Effect—an effect different from Arrow's impossibility theorem because Condorcet clearly stated that the question of the vote is not to find a device that avoids hurting some private interests, but to find a solution complying with the general interest even at the expense of the interests of some private individuals (Faccarello 2016b).

But even if the Legislator makes the right decisions, Turgot insisted, one of the main obstacles to a policy aimed at the public interest is all kinds of private interests embedded in public opinion: this is the reason why they must be dealt with carefully, to avoid violent reactions and putting the country upside down: "one must ... only work slowly, as public opinion and the course of events make these changes possible" (Turgot 1913–23, V, p. 184). But this does not mean that the enlightened Legislator must remain passive. Public opinion is a sort of unavoidable interlocutor or a kind of tribunal before which public affairs had to be debated in order to convince it. But what is this public? Turgot knew that the public was not a homogeneous entity. In a nutshell, he distinguished three different kinds of people on three levels of the

social scale. At the bottom of this scale are the “people”, who are poor, ignorant and suspicious. In order to be convinced, they must “to reflect and ... compare the future with the present”, but are incapable of doing so because of their “habit of living from day to day” and “the kind of apathy” into which poverty plunges them (Turgot 1913–23, II, p. 235). It is not the people who must be addressed, at least as a priority. At the other end of the social scale are the privileged. These persons are generally well-educated, but caste prejudices and short-term interests often make their judgment bad—in this category are also included all those, who, belonging to intermediary bodies in society (like the members of the Parliaments, but also the corporations, etc.), tend to make the private interests of these bodies prevail over the general interest. What is the use therefore to trying to convince them as a group? This is a waste of time and energy—Turgot responded with the most absolute scepticism to Dupont, who wrote a tragedy with the hope of persuading princes and all those in power by spreading Physiocratic ideas in an attractive and didactic form. Finally, in the middle of the social scale are persons whose position shelters them from an incessant concern with basic needs, and who have acquired the habit of becoming informed, reading, and discussing: in a word, they use their reason. They form the enlightened class of the population. It is their opinion that matters, for they are the ones who form the true public: “it is to the reading and thinking public that we must speak; they are the ones whom we must please and the only ones we need persuade” (Turgot 1913–23, III, p. 491)—hence the importance, not only that this part of the population be the largest possible, but also that it be the best instructed. To enlighten the public means in the long run to propagate the ideas of the Enlightenment both towards the bottom and especially towards the top of the social scale. For what do princes who do not use their reason do, if not follow the dominant opinion? They “will truly be convinced only of what everyone surrounding them is convinced of, i.e. truths that have become ... popular” (Turgot 1913–23, III, p. 481): “it is the only means of lifting that opposition that is born on all sides by misunderstood interests” (Turgot 1913–23, II, p. 515). Baudeau was on the same opinion (in Turgot 1913–23, IV, p. 222).

Say’s views on the Legislator put great emphasis on the role of the administration in devising economic policies that promote the general interest. Following the views of the French Enlightenment with its emphasis on education, he considered that political economy was necessary for the alignment of interests within the nation. Political economy is a useful and practical science for the citizens—for example, those entrepreneurs who mistakenly ask for protective duties—who can thus understand better their own self-interest and the functioning of the society itself. This is the explicit goal of his *Cours complet d’économie politique pratique*, as mentioned in the long subtitle: “book made for showing the economy of the society to statesmen, landlords and capitalists, scientists, farmers, manufacturers, merchants and, in general, all the citizens”. The political dimension of Say’s “philosophie économique” aims at producing a knowledge that can be spread among the public opinion, so that an enlightened opinion could check and control the decisions of the government, notably through the vote of the budget.

An enlightened administration is also needed to implement the policies enacted by wise rulers. It is important to bear in mind the fact that, in his *Traité*, Say's "Discours préliminaire" ends with a long discussion of the respective role of public opinion, the legislator, the administration and the teaching of political economy (Say 1825, I, pp. xcv–cviii). The administration is, according to Say, a key element in the sense that the implementation of a good policy requires an administration able to understand the effects of the policy at stake. Without such administrative skills, the legislator and the citizens would be deceived in their attempts to improve the economy of the society: "And although a monarch and his principal ministers should be well acquainted with the principles upon which national prosperity is founded, of what advantage this knowledge be to them, if throughout all the departments of administration, their measures were not supported by men capable of comprehending and enforcing them? The prosperity of a city or province is sometimes dependent upon the official acts of a single individual and the head of a subordinate department of government, by provoking an important decision, often exercises an influence superior to that of the legislator himself" (Say 1825, I, pp. xcv–xcvi). But beyond this important point, the role of the citizens themselves remain fundamental: "In countries blessed with a representative form of government, each citizen is under a much greater obligation to make himself acquainted with the principles of political economy; for there every man is called upon to deliberate upon public affairs ... A nation, in order to enjoy the advantage of a good system of political economy, must not only possess statesmen capable of adopting the best plans, but the population must be in a situation to admit it of their application" (Say 1825, I, pp. xcvi–xcvii). Beyond these general considerations, in the then unpublished manuscript of his "Politique pratique" (practical politics), Say asked: "Is it possible to benefit from the principle of competition, a principle so fruitful, in practical politics? Yes: and by the same token, it is possible to benefit from the variety of the human nature, as well very powerful. To benefit from a vice is the ultimate moral skillfulness" (Say 2003, pp. 432–433). The principle of competition is called here "principle of emulation". The political arrangement Say had then in mind rests on the existence of two assemblies, the white and the blue, elected in the same way: every time one assembly proposes a new law, the other one examines the law and put its "vanity" to find any "imperfection" in the proposed act as a reason for rejecting it—alternatively, it can also accept it for fear of the public outrage that would follow a rejection whenever the alleged defects are only slight imperfections (Say 2003, p. 433). The idea was clearly to make vanity work for the general interest in the domain of politics, just as competition did with self-interest in the economy.

## 4 Conclusion

Over the long period considered here, that is, from 1695 to 1830, the authors belonging to the "Philosophie économique" current of thought were active in promoting the social benefits that a society and the government would get from



self-interest *cum* competition. Relentlessly, they fought against alternative views—such as the “moral economy” or the “military economy” approaches—in order to convince public opinion that radical changes were to be made in order to reach a better economic and political situation. But was it possible to extend the realm of self-interest as far as to make it the basis of the whole social fabric, at a time of the first economic crises and emergence of pauperism?

In the 1830s, Say, the last important member of this current of thought, was confronted to new and numerous opponents, who rejected the views according to which other spring of actions than self-interest and competition were unimportant for the smooth functioning of the new industrial society. Religious beliefs (Faccarello and Steiner 2008b; Faccarello 2017) or altruistic behaviours (Steiner 2017) were brought to the fore by liberal political philosophers (Germaine de Stael, Benjamin Constant) and various religious or socialist thinkers (Alban de Villeneuve-Bargemont, Charles de Coux, Pierre Leroux), by those claiming that a new religion was required (Henri Saint-Simon) or by those endorsing the idea that another “new science” was necessary (Auguste Comte) to stabilise a society plagued with pauperism and crises. “Philosophie économique” faded away, but the centrality of the self-interested behaviour is still with us.

## References

- Baudeau, N. 1771. *Première introduction à la philosophie économique, ou Analyse des États policés, par un disciple de l'Ami des Hommes*. Paris: Didot, Delalain & Lacombe.
- Boisguilbert, P. Le Pesant de. 1966. ‘Correspondance’ and ‘Œuvres. In *Pierre de Boisguilbert ou la naissance de l'économie politique*, ed. J. Hecht. Paris: INED.
- Charles, L., and Ph. Steiner. 1999. Entre Montesquieu et Rousseau: la Physiocratie aux origines de la Révolution française. *Études Jean-Jacques Rousseau* 11: 83–159.
- Charles, L., F. Lefebvre, and C. Théré, eds. 2011. *Le cercle de Vincent de Gournay. Savoirs économiques et pratiques administratives en France au milieu du 18e siècle*. Paris: INED.
- Condorcet, M.-J.-A.-N. Caritat (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris: Imprimerie Royale. (The important ‘Discours préliminaire’ is reproduced in M.-J.-A.-N. C. de Condorcet (1986), *Sur les élections et autres textes*, Paris: Fayard, 9–177.
- Condorcet, M.-J.-A.-N. Caritat 1847–49. *Œuvres*, A. Condorcet-O'Connor and F. Arago, ed., 12 vols. Paris: Firmin Didot.
- Condorcet, M.-J.-A.-N. Caritat 1994. *Condorcet. Arithmétique politique: textes rares ou inédits (1767-1789)*, ed. with comments by B. Bru and P. Crépel. Paris: INED.
- Demals, Th., and G. Faccarello. 2016. French enlightenment. In *Handbook on the history of economic analysis*, Faccarello, G., and H. D. Kurz, ed., vol. II, *Schools of Thought in Economics*, Cheltenham: Edward Elgar, 75–91.
- Faccarello, G. 1986. *Aux origines de l'économie politique libérale: Pierre de Boisguilbert*, Paris: Anthropos, revised edn, *The Foundations of Laissez-faire. The Economics of Pierre de Boisguilbert*, London: Routledge, 1999.
- Faccarello, G. 1998. Galiani, Necker and Turgot: A debate on economic reform and policy in 18th century France. In *Studies in the history of French Political Economy. From Bodin to Walras*, ed. G. Faccarello, London: Routledge, 120–185.



- Faccarello, G. 2006. An 'exception culturelle'? French Sensationist political economy and the shaping of public economics. *The European Journal of the History of Economic Thought* 13 (1): 1–38.
- Faccarello, G. 2010. Bold ideas. French liberal economists and the State: Say to Leroy-Beaulieu. *The European Journal of the History of Economic Thought* 17 (4): 719–758.
- Faccarello, G. 2016a. Anne-Robert-Jacques Turgot (1727-1781), *Handbook on the history of economic analysis*, Faccarello, G., and H. D. Kurz, eds., vol. I. *Great Economists since Petty and Boisguilbert*, Cheltenham: Edward Elgar, 73–82.
- Faccarello, G. 2016b. Marie-Jean-Antoine-Nicolas de Condorcet (1743–1794). In *Handbook on the history of economic analysis*, vol. I, Faccarello, G., and H. D. Kurz, ed. *Great Economists since Petty and Boisguilbert*, Cheltenham: Edward Elgar, pp. 83-94.
- Faccarello, G. 2017. A dance teacher for paralysed people? Charles de Coux and the dream of a Christian political economy. *The European Journal of the History of Economic Thought*, 24(4), 828–875.
- Faccarello, G., and Ph. Steiner. 2008a. Interest, sensationism and the science of the legislator. French 'Philosophie économique', 1695–1830. *The European Journal of the History of Economic Thought*, 15(1), 1–23.
- Faccarello, G., and Ph. Steiner. 2008b. Political economy and religion in early nineteenth century France. In *Keeping Faith, Losing Faith: Religious Belief and Political Economy*, Bradley W. Bateman and H. Spencer Banzhaf, eds. Annual supplement of *History of Political Economy*, 40, 26–61.
- Le Mercier, de la Rivière, P.-P. 1767. *L'ordre naturel et essentiel des sociétés politiques*. London: Jean Nourse, and Paris: Desaint.
- Le Trosne, G.-F. 1788. *De l'administration provinciale et de la réforme de l'impôt*. Bâle and Paris: Duplain.
- Mirabeau, V.-Riqueti. 1761. *Théorie de l'impôt* [n.p.].
- Mirabeau, V.-Riqueti. 1769. *Lettres sur la dépravation de l'ordre légal*. Amsterdam: J. H. Schneider.
- Morellet, A. 1790. Mémoire lu au Comité des contributions publiques. *Mémoires d'économie publique, de morale et de politique*, An VIII (1800), 2(2), 100–119.
- Nicole, P. 1670. *De l'éducation d'un prince*. Paris: Veuve Charles Savreux.
- Oudin-Bastide, C., and Ph Steiner. 2019. *Calculation and morality in the French debate over the abolition of slavery*. Oxford: Oxford University Press.
- Quesnay, F. 2005. *Œuvres économiques complètes et autres textes*, C. Théré, L. Charles and J.-C. Perrot, eds. Paris: INED.
- Rieucan, N. 1998. 'Les entreprises où les hommes s'exposent à une perte dans la vue d'un profit': Condorcet et l'héritage de d'Alembert. *Revue économique* 49 (5): 1365–1405.
- Say, J.-B. 1803. *Traité d'économie politique*, 1st ed. Paris: Deterville.
- Say, J.-B. 1817. *Petit volume contenant quelques aperçus des hommes et de la société*, Paris: Deterville.
- Say, J.-B. 1825. *Traité d'économie politique*. Fifth edition, Paris: Rapilly.
- Say, J.-B. 1828–29. *Cours complet d'économie politique pratique*. Paris: Rapilly.
- Say, J.-B. 2003. *Politique pratique*. In J.-B. Say, *Œuvres complètes*, vol. 5, *Œuvres morales et politiques*, Paris: Economica, 287–822.
- Steiner, Ph. 1994. Demand, price and net product in the early writings of François Quesnay. *The European Journal of the History of Economic Thought* 1 (2): 231–251.
- Steiner, Ph. 2006. Say et le libéralisme économique. In *Histoire du libéralisme en Europe*, ed. P. Nemo and J. Petitot, 381–403. Paris: Presses universitaires de France, 381–403.
- Steiner, Ph. 2014. Administration and economic government in Quesnay's political economy. In *Economics and other branches—In the shade of the oak tree*, ed. R. Baranzini and F. Allisson, 123–132. London: Pickering and Chatto, 123–132.
- Steiner, Ph. 2017. Religion and the sociological critique of political economy: altruism and gift. *The European Journal of the History of Economic Thought* 24 (4): 876–906.

- Théré, C. 1998. Economic publishing and authors: 1566-1789. In *Studies in the history of French political economy, From Bodin to Walras*, ed. G. Faccarello, 1–56.
- Thompson, E.P. 1971. The moral economy of the English crowd in the Eighteenth century. *Past and Present* 50 (February): 76–136.
- Turgot, A.-R.-J. 1913–23. *Œuvres de Turgot et documents le concernant*, Gustave Schelle, ed., 5 vols, Paris: Félix Alcan.

# Self-interest in the Thought of Adam Ferguson



Craig Smith

**Abstract** Adam Ferguson (1723–1816) was a prominent member of the Scottish Enlightenment. His most famous work *An Essay on the History of Civil Society* (1767) has often been read as a dissenting voice from the positive view of commerce found in the work of his friends Adam Smith and David Hume. Readings of Ferguson tend to focus on the *Essay* and to see him as either a civic republican worried about the impact of commerce on citizenship, or as a precursor to Marxian ideas of alienation and the anti-social impact of economic development. This paper argues against both of these interpretations, and against the practice of reading Ferguson through the *Essay* alone. Taking his discussion of self-interest as its focus, the paper shows how, in Ferguson's other writings, he develops a complex and nuanced understanding of the place of self-interest in moral and political life. Central to this is Ferguson's concept of ambition: an idea crucial to his moral philosophy and one which places Ferguson at the heart of eighteenth century debates about the nature of self-regarding behaviour.

**Keywords** Adam Ferguson · An Essay on the History of Civil Society · Moral philosophy · Self-Interest · Ambition

## 1 Introduction

There has been a pronounced revival of interest in the thought of Adam Ferguson over the last thirty years. Once seen as a marginal figure in the history of ideas, he has now been recognised as an original voice within the Scottish Enlightenment. Ferguson scholarship has developed along a number of well-established lines: he has been read as a late entry in the civic tradition whose old-fashioned moralism stands in contrast to his friends David Hume and Adam Smith's more sanguine embrace of commercial modernity, and as an early figure in the development of sociology

---

C. Smith (✉)

School of Social and Political Sciences, The University of Glasgow, Adam Smith Building, 40 Bute Gardens, Glasgow G12 8RT, UK  
e-mail: [Craig.Smith@glasgow.ac.uk](mailto:Craig.Smith@glasgow.ac.uk)

whose theory of socialisation leads him to a critical perspective on the impact of the division of labour that prefigures later ideas of alienation.<sup>1</sup>

In both cases Ferguson is read as a significant critic of commercial modernity and as an opponent of the pursuit of self-interest at the expense of the community. There are good reasons for both of these interpretations, and good reasons for them to stress the sceptical passages in Ferguson's work. The scepticism about commercial society is most obvious in the highly wrought rhetoric of Ferguson's most famous work the *Essay on the History of Civil Society*.<sup>2</sup> Here we find Ferguson's famous worry that pursuit of self-interest will lead us to 'break the bands of society' and 'dismember the human character' (1995: 218).

The two schools of interpretation of Ferguson dwell on these particular images and what they say about his attitude to commerce, but they do so for quite different reasons. The sociological reading sees these passages as evidence that Ferguson had developed a proto-alienation theory. That the division of labour and commerce was somehow mutilating and inhuman and that it was commerce that separated the individual from the community. The civic reading, on the other hand, views the dismemberment in political terms. It sees Ferguson as concerned with the loss of an active citizenry as attention is drawn to commerce and luxury. In particular it contends that he was worried that the separation of the roles of statesman and warrior would leave an emasculated population at the prey of its enemies and of unscrupulous populists and demagogues.

Ferguson's life-long obsession with the creation of a Scottish militia indicates that he is drawing on the civic tradition to highlight the danger of a professional political class and an effeminate and supine populace, unwilling and unable to act politically to defend the nation. As the division of labour advances, so the analysis goes, we become more limited in the scope of our skills and at the same time more self-absorbed. As he puts it: 'Where shall we find the talents which are fit to act with men in a collective body, if we break that body into parts, and confine the observation of each to a separate track?' (Ferguson 1995: 32).

Both the civic and the sociological readings of Ferguson see him as a critic of the moral effects of selfishness and the corruption of the human character that comes from the pursuit of luxury in a commercial society. The apparent scepticism about commerce and the moralistic criticism of self-interest has led to Ferguson being

---

<sup>1</sup>The exploration of the Civic Republican tradition by Pocock (1975) has set the agenda for much work in the history of political thought of the early modern period. Classic civic or republican readings of Ferguson are Robertson (1985), Sher (1989), and Sebastiani (2011; 2013: 50). Berry (2013: 154) argues that Ferguson is sympathetic to republicanism but does not subscribe to it in detail as he recognises that the conditions of modernity preclude it. McDaniel (2013a: 11) goes further and argues that Ferguson is actually a critic of republicanism. Sociological readings of Ferguson include MacRae (1969), Swingewood (1970), Benton (1990), Kettler (2005), and Brewer (1986; 1989; 2007). Forbes (1982: 193) criticises the sociological readings. For discussions of Ferguson's relationship to the Marxian tradition and the concept of alienation see (Mizuta 1981; Meek 1967; Hamowy 1968; Skinner 1982; Brewer 1986; Benton 1990). Lisa Hill (2001, 2007) and Jack Hill (2013) are sceptical of the Marxist reading of Ferguson.

<sup>2</sup>For example, J.G.A. Pocock's (1975; 1999) discussions of Ferguson focus almost entirely on the *Essay*.

used as a foil for Hume and Smith in discussions of the Scottish Enlightenment's attitude to commercial society.<sup>3</sup> Ferguson's views on economics have thus been characterised as the last gasp of the civic republican tradition or the first breaths of the communitarian or Marxist critique of commercial modernity. Framing the writing of the period as a debate between the old-fashioned moralist of the civic tradition or an anticipator of the Marxian tradition, and the proponents of commercial liberty is a neat organisational device, but it has provided us with a distorted picture of the thought of the time and of Ferguson's place in it. Nowhere is this more obvious than in the perception of what Ferguson is supposed to have thought about self-interest.

This distorted reading of Ferguson stems chiefly from him being read almost exclusively through the *Essay* and with one eye on his interlocutors. If, instead of this, we focus on what Ferguson argues across all of his publications, a very different view of his understanding of self-interest emerges.<sup>4</sup> From the early lecture headings of the *Analysis of Pneumatics and Moral Philosophy* (1766) to the final published version of his lectures in the *Principles of Moral and Political Science* (1792) Ferguson develops a complex and nuanced account of self-interest and self-regarding behaviour. To help us understand this it is helpful to set out a more appropriate context for the development of Ferguson's thinking on these matters.

## 2 Context

The context for this discussion is the systematic account of moral philosophy that Ferguson developed during his time as the Professor of Pneumatics and Moral Philosophy at the University of Edinburgh. Ferguson enters a debate in Scottish moral philosophy that was inspired by the challenge of Bernard Mandeville's satirical account of selfishness and benevolence in *The Fable of the Bees*. Mandeville's famous observation that the 'Moral Virtues are the Political Offspring which Flattery begot upon Pride.' (Mandeville 1988 Vol. 1: 51) allowed him to produce a theory which saw humans as fundamentally selfish creatures who are manipulated into society and into believing in moral principles by an elite of cynical politicians.<sup>5</sup> At its most extreme, Mandeville's account denies the possibility of benevolence. Take the following example from his contribution to the *Female Tatler*:

...if the most publick spirited Man in the Universe will be pleased strictly to examine himself, he will find that he has never committed any action deliberately but for his own sake; for he

---

<sup>3</sup>Examples of Ferguson being used as a foil for Hume and Smith include: Winch (1978: 175), McDowell (1983), Sher (1985: 237; 1989; 1994: 394-8), Allan (2006: 68), Graham (2013: 512-13), and Arbo (2011).

<sup>4</sup>Some other scholars have made similar attempts to read Ferguson within his own system of thought, most prominently David Kettler (2005), Lisa Hill (2006), and Jack Hill (2017). My own recent book takes a similar approach (Smith 2018).

<sup>5</sup>M. M. Goldsmith describes Mandeville as arguing that: 'Society and morality are set up by clever, selfish, vicious, cynical superior beings manipulating selfish, vicious, but susceptible and gullible, inferiors. It is a trick played on fools by knaves.' (Goldsmith 1985: 53).

had a satisfaction either in the doing of it or in the hopes of Praise, which sooner or later he wou'd receive for it (Mandeville 1999: 152).

Mandeville lingers behind much of the social theory of the Scottish Enlightenment, but he has particular prominence in the work of Francis Hutcheson. Hutcheson penned a direct response to Mandeville, *Remarks upon the Fable of the Bees* (1750), and throughout his works set himself the task of undermining the reduction of morality to self-interest.<sup>6</sup> His chief arguments for this were to demonstrate that self-interest could not explain many ordinarily experienced elements of moral life, and that, on the contrary, a realistic account of morality recognised the reality of benevolence. The point at which Hutcheson pries open Mandeville's argument is the account of apparently benevolent actions being, in fact, the result of a desire for praise arising from human pride. While this argument had plenty of shock value for Mandeville on account of its counterintuitive cleverness, it also left him open to obvious counter-examples that Hutcheson uses to convince his readers of the weakness of Mandeville's argument on benevolence.

Hutcheson's own system regarded benevolence as the central principle of moral experience, and the key feature guiding our moral sense. But he also realised something else: that Mandeville's argument was plausible precisely because selfishness did appear to be a significant motivation in many areas of human life. Moreover, Mandeville had shown how, in some cases, 'Private vices by the dextrous Management of a skilful politician may be turned into Publick Benefits' (Mandeville 1988 Vol. 1: 369). The question that faced Hutcheson, and indeed those who followed him in the Scottish Enlightenment, was what is the proper relationship between benevolent and self-interested behaviour in a commercial society?

In a sense this is the problem that lies behind David Hume's defence of commerce in his essays on political economy, and even more prominently of Adam Smith's account of the moral sentiments, which he famously begins with the line:

How selfish soever mankind may be supposed, there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it (Smith 1976: 9).

Smith includes several passages in *The Theory of Moral Sentiments* where he attacks the 'selfish' or 'licentious' system of Mandeville (Smith 1976: 312–13). But early on in his career he made a wider observation on the context of Mandeville's argument. In his *Letter to the Edinburgh Review*, Smith (1980: 250) notes the similarity between Rousseau's account from the *Discours sur l'origine et les fondemens de l'inégalité parmi les hommes* (1755), and Mandeville's argument. The idea that there might be benign and malign versions of self-regarding behaviour became an obvious response to the challenge of Mandeville. For Smith the answer lay in a careful analysis of the different elements of moral experience. There was, Smith thought, some explanatory mileage in the jurisprudential distinctions between justice

---

<sup>6</sup>See, for example (Hutcheson 2002: 29; 2004: 92–5). For a discussion of Mandeville's concept of selfishness see Monro (1975) and Hundert (1994), for a discussion of his reception in the Scottish Enlightenment see Goldsmith (1988).

and benevolence that would allow an explanation of the different degrees of appropriateness of self-regarding behaviour. One may indeed be just by sitting still and doing nothing but one would be open to the resentment of one's peers if such inaction displayed a lack of appropriate beneficence (Smith 1976: 82).

His other way of approaching this was by recognising that prudence was indeed a virtue, and one that must be reconciled with the other elements of moral experience. Smith clearly thought that his own theory of the moral sentiments provided an accurate account of the appropriate place of self-interest in the grand scheme of moral life. It's worth noting though that Smith's two main contemporary critics press him precisely on the relationship between selfish and other regarding sentiments. David Hume wrote to Smith after the first edition of the *Theory* and addressed him on the point that he had taken insufficient care to illustrate that if all kinds of sympathy are agreeable he ran the risk of implying that we felt pleasure alongside unpleasant emotions (Smith 1987: 47).<sup>7</sup> And his successor at Glasgow, Thomas Reid, dismissed his system as a refinement of the selfish system (Reeder 1997: 77, 80–81) precisely because the whole depended upon the pleasure felt alongside the moral sentiments. From all of this it should be obvious that the balance between the self and other regarding elements of human behaviour was central to the moral philosophy of the period.

### 3 Ferguson

The immediate philosophical context for Ferguson's discussion of self-interest is this heady mixture of attempts to square the apparent contradiction between benevolence and self-interest as motivations for morally acceptable behaviour. Like Francis Hutcheson, Ferguson wants to defend the reality of benevolence and place it at the heart of his normative theory. Mandeville, according to Ferguson, is wrong because: 'It is ridiculous to attempt confounding the love of virtue with pride: their tendencies are opposite.' (Ferguson 1994: 104).

Ferguson's criticism of Mandeville (1994: 104; 1995: 18–20, 36–8; 2006: 213, 250) is based on the idea that the literary effect of deliberately subverting the ideas of vice and virtue, in reality reveals the solidity of the concepts. Mandeville's reversal is daring and witty precisely because it goes against the evidence of human experience. The evidence suggests that 'expressions of praise and blame in every language...show, that the distinction of moral good and evil is real, and universally acknowledged.' (Ferguson 1994: 108). But in making this move Ferguson has learned the lesson of Mandeville's rigorous attempt to regard virtue as purely selfless action. A convincing moral philosophy must provide a convincing account of the actual relationship between self-regarding and other-regarding behaviour.

---

<sup>7</sup>David Hume was famously dismissive of the *Essay*, for a discussion of which see Merolle (2009) and Raynor (2009).

In order to find this Ferguson turns to his established methodology of empirical investigation. He believed that there was a universal human nature and that the apparent diversity of customs and habits was underwritten by a core set of universal passions: 'They engage in different pursuits, or acquiesce in different conditions; but act from passions nearly the same,' (Ferguson 1995: 51). He also believed that underneath the apparent diversity of human institutions and beliefs lay a set of commonly experienced sets of circumstances with which every society had to contend. By identifying these underlying passions and circumstances Ferguson hoped to create a reliable empirically justified basis for the practice of normative moral philosophy. By understanding the nature of Man we will be able, like Aristotle, to develop moral principles that should guide his behaviour.

Ferguson's account of the 'Principle of Society in human nature' (1973 Vol. 1: 26) treats it as a 'fact' (1995: 39) supported by 'irrefragable proofs' (1973 Vol. 1: 47) and established by universal observation. His account of the evolution of human sociability is particularly sophisticated, and the basis for his reputation as a founding father of sociology. It is not calculated self-interest that first attaches us to society for Ferguson for the very good reason that humans are social before they are rational. Ferguson suggests that society is the source of 'almost the whole of his rational character.' (1995: 23). This means that there is a fundamental error in both Mandeville's account and in Social Contract theories and their accounts of the state of nature. They describe a situation before society in their states of nature, a situation that Ferguson believes cannot be supported by the evidence and which he calls a 'mere fiction in theory' (1994: 220).<sup>8</sup> They then assume that a rational calculation and choice is made to enter into society or submit to government. Ferguson notes that we have no evidence of this, but we do have ample evidence for the pleasure that humans take in social life. Humans are miserable if they are cut off from human contact and their happiness depends on interaction with others. So, for Ferguson, humans come already attached to other humans and the choice to associate is not conscious or interested, but rather an expression of our nature as social animals. Man is 'formed for society, and is excellent in the degree in which he possesses the qualifications of an associate and a friend.' (1973 Vol. 2: 41).

But Ferguson goes further than this, morality itself is a social phenomenon, and its content is dictated by humanity's social disposition.

The great distinction of right and wrong, of virtue and vice, on which men experience such extremes of complacency or indignation, of esteem or contempt, is formed on the dictates of a social disposition, which receives, with favour and love, what constitutes the good of mankind, or rejects, with disapprobation and abhorrence, what is of a contrary nature (1973 Vol. 1: 35).

This is an argument that runs through Ferguson's writings from his first publication in moral philosophy the *Analysis* where 'Man is by nature the member of a community. His love of that community renders him a good member of it, and intitles him to praise.' (1766: 32), to his final book the *Principles* where:

---

<sup>8</sup>For his attack on social contract theory see *Institutes* (1994: 219-225) the *Essay* (1995: 8-15, 73), and the *Principles* (1973 Vol. 1: 195-99) (1973 Vol. 2: 460).



Society, in which alone the distinction of right and wrong is exemplified, may be considered as the garden of God, in which the tree of knowledge of good and evil is planted; and in which men are destined to distinguish, and to chuse, among its fruits (1973 Vol. 1: 268).

The act of moral judgement is a social act in Ferguson's view. What he calls 'censorial inspection' (1973 Vol. 2: 22) is undertaken within a community and is based on universal traits of human nature. As he puts it:

The distinction of right and wrong is coeval with human nature: It is perceived without instruction, in acts of fidelity and beneficence, or of perfidy and malice. These are the topics of praise and blame, in every nation and in every age. That, indeed, which in one instance is considered a benefit, in another instance is considered as harm or detriment (1973 Vol. 1: 300).

On a basic level this leads him to observe that it is a universal principle of human moral behaviour that benevolence is regarded as pleasant while malice is regarded as painful (Ferguson 1994: 179). It is a basic datum for Ferguson that humans are social. They are 'formed for society' (Ferguson 1973 Vol. 2: 41), and this fact of their nature explains why benevolence must become a central part of any attempt to understand their moral beliefs. For Ferguson, benevolence is the 'specific excellence of man' (Ferguson 1973 Vol. 2: 34), and the 'greatest good to which human nature is competent' (Ferguson 1973 Vol. 2: 149), so in Ferguson's system this 'love of mankind' (Ferguson 1994: 116) provides the content for morality.<sup>9</sup>

But again, we need evidence for this, and evidence that refutes the alternative account, that we are social for the selfish reason that we need protection from the group or that we profit from social interaction. Ferguson believes that the decisive evidence against such accounts lies in our willingness to fight to defend our group. As he puts it: 'Men are so far from valuing society on account of its mere external conveniencies, that they are commonly most attached where those conveniencies are least frequent.' (Ferguson 1995: 23). Ferguson's is not a crude and cosmopolitan notion of benevolence, but rather an account grounded in the evidence of the actual experience of sociability by individuals who are embedded in a nation. One facet of this, and one that forms part of Ferguson's reputation as a founding father of sociology, is his account of the generation of national identity through an early account of conflict sociology.

For Ferguson humans assemble in 'troops and companies' (1995: 21) and their identity is shaped as part of larger social groups. This means that self-interest arises, historically, in a social setting and always involves a connection to the group. We identify with the group and do so in opposition to other groups. This provides a far more persuasive psychology to the act of dying in war than Mandeville's convoluted

---

<sup>9</sup>Though we should be aware that Ferguson hedges his identification of benevolence as the supreme principle of morality: 'We are thus disposed, for the most part, to simplify our conceptions, and to seek for some fundamental principle from which, if we secure it, every other requisite will follow; but the reality of any such principle, even in the government of human nature, may be doubted. The love of mankind, which we have assumed as the nearest approach to a general principle of virtue, requires the direction of wisdom and the support of courage.' (1973 Vol. 2: 332-3).

deceived pride account.<sup>10</sup> Giving up one's life in defence of the nation is an act of benevolence – though crucially not one of cosmopolitan benevolence.

## 4 Ambition

This empirical approach also leads Ferguson to recognise the reality of self-regarding behaviour by individuals within the nation. In fact, far from backing away from the idea of self-interest as we might suppose given his statements on benevolence and the contemporary readings of him as sceptical about commerce, he instead embraces it as perhaps the central concept of his account of moral behaviour. Ferguson does this by way of his account of ambition. Ambition is the 'desire of something better than is possessed at present.' (Ferguson 1973 Vol. 1: 202), and it is this that prompts human activity. Without ambition there is no prompt to action.<sup>11</sup>

if the desire of any thing better than the present should at any time cease to operate in his mind, he becomes listless and negligent, loses the advantages he had gained, whether of possession or skill, and declines in his fortune, till a sense of his own defects and his sufferings restore his industry (Ferguson 1973 Vol. 1: 56).

How, then, does Ferguson manage to square the account of benevolence with an account of human motivation to action grounded in ambition? The answer is that he recognises that ambition is ubiquitous, that it motivates us in both the material, economic world and in the moral world. In short, we are ambitious to be better people.

Ferguson traces this to the universal human propensity to make judgements. We judge one material condition to be preferable to another, and we judge one form of action to be better than another. Ferguson is able to do this because of his account of choice and discernment. Humans make qualitative judgements: we choose what we think is better. Humans are able to form a conception of something better than they have at present, and the difference between this ideal and reality is what urges us to improvement.<sup>12</sup>

If we are asked, therefore, what is the principle of moral approbation, we may answer, It is the *Idea of perfection* or excellence, which the intelligent and associated being forms to himself; and to which he refers in every sentiment of esteem or contempt, and in every expression of commendation or censure (Ferguson 1973 Vol. 2: 134).

---

<sup>10</sup>For an extended version of Mandeville's account of the deception of honour as an explanation for the willingness to die for our country see his *An Enquiry into the Origin of Honour and the usefulness of Christianity in War* (1732).

<sup>11</sup>'Ambition or the desire of Elevation or Perfection is certainly among the Strongest or the most powerful [in] the human Breast.' (Ferguson 2006: 80).

<sup>12</sup>Lisa Hill (2006: 97) thinks that Ferguson attempts to distinguish between ambition and interest: 'Interest is largely confined to the care of the body whereas ambition is generally directed towards progress, whether moral, intellectual or technical.' But there is little evidence for such a systematic distinction in Ferguson's account. Ambition is used indiscriminately to apply to both material and moral improvement.

The recognition that we have yet to achieve the ideal is what prompts us to act. This is as true for economic activity as it is for our attempts to improve as moral agents. Recognition that our situation is not optimal is necessary for there to be improvement. As he puts it: ‘Complaints of moral evil are the symptoms of a progressive or improving nature. A being that perceived no moral evil, or no defect, could have no principle of improvement.’ (Ferguson 1994: 132).

The account of moral ambition is a central part of Ferguson’s thought throughout his career. The account of moral judgement, or what he calls ‘censorial inspection,’ is certainly an account of judging others, but it is also an account of self-judgement. When Ferguson develops his account of moral responsibility it is based on the fact that we are self-aware and capable of assessing ourselves against ideas of perfection. As he would have it: ‘It is a privilege of the mind to contemplate itself; to chuse its inclinations and sentiments; to suppress what is evil, and cultivate what is good.’ (Ferguson 1766: 23).

Human ambition in the moral sphere is the ambition to be better people, to act in line with the conception that we form of our better selves. The difference between this and material ambition is not one of kind, but rather one of object. So far, so good. Ambition as the general propensity behind all human activity applies to both moral and commercial activity. But what happens when the ambition to be morally better comes up against the ambition to be materially better off? One way to understand Ferguson’s position on this is to look at the passages where he discusses what he calls the ‘laws’ of ‘self-preservation’ and of ‘society’ (Ferguson 1994: 91). Ferguson is clear that these are not contradictory. We desire both our own welfare and the welfare of others and we are ambitious to achieve both of these. The two are not inconsistent: ‘If the law of self-preservation, for the most part, prevail, it does not follow, that the law of society has no effect.’ (Ferguson 1994: 92).

This leaves us with the question of how Ferguson believes he can bring the two together. In the *Institutes* he answers this by accepting Adam Smith’s point that existing systems of philosophy have identified self-interest and benevolence as distinct principles and presented us with a choice: either all moral behaviour is descended from benevolence or it is descended from self-interest. But this, according to both Smith and Ferguson, is a false dichotomy, a forced choice that need not be made and that does not capture the observed reality of human life. As Ferguson puts it: ‘the laws of self-preservation, and of society, when well understood, coincide in all their tendencies and applications.’ (Ferguson 1994: 115).

So far from acting as a foil for Smith on this issue he is explicitly agreeing with his position. As Donald Rutherford (2012: 259-60) has shown, Ferguson’s account of the benefits of action pays close attention to the incentives that prompt activity. In the economic sphere this means that interest is sufficient to move us to seek to improve our material condition and to provide for our family. Ferguson repeatedly argues that these incentives alone will drive forward economic growth and lead to a rising population. Governments should have no part in trying deliberately to bring about these ends as:

...communities become rich, not from the impulse of public institutions, but rather from the ambition of their separate members, who wish to provide for themselves what is considered as a constituent of superiority in the distinction of ranks (Ferguson 1973 Vol. 1: 237).

It would appear that Ferguson's concept of ambition intimately connects material and moral activity so the idea that Ferguson is setting self-interest in contradiction to benevolence simply will not stand up. The account of corruption and the worries about commercial society then, are not what they have often taken to have been. We find further support for this if we focus on Ferguson's discussion of luxury in the *Essay* itself: 'The *necessary of life* is a vague and a relative term: it is one thing in the opinion of the savage; another in that of the polished citizen: it has reference to the fancy, and to the habits of living.' (Ferguson 1995: 137-8).

Ferguson is able to make this point because he has a particular understanding of what motivates humans when it comes to wealth: 'Riches consist in the abundance of things that conduce to safety, subsistence, accommodation, and ornament.' (Ferguson 1994: 31). And these objects are pursued in at every level of economic development. They are coeval in human behaviour, what Christopher Berry (2009) calls Ferguson's 'principle of simultaneity.'

The convenient and ornamental in their several forms, however rude, are studied in the same age with the necessary; and the same person, who subsists from meal to meal or the precarious returns of the chace, is, in the intervals of his necessity, no less studious of ornament in his person, his dress, and the fabric of his habitation, his weapons, or arms, than he is earnest in procuring his food (Ferguson 1973 Vol. 1: 240).

For all of the rhetoric about breaking the bonds of society Ferguson could not be clearer that luxury is a relative concept. A cottage is a luxury to one who lives in a cave (Ferguson 1995: 12–13).

Ferguson's point is exactly the same as that made by Hume in his attack on vicious luxury in the essay *Of Refinement in Arts* (Hume 1985: 279). He is quite clear about this even at his most sceptical in the *Essay*: 'The use of morality on this subject, is not to limit men to any particular species of lodging, diet, or cloaths; but to prevent them considering these conveniencies as the principal objects of human life.' (Ferguson 1995: 234). And again: 'It appears, therefore, that although the mere use of materials which constitute luxury, may be distinguished from actual vice; yet nations under a high state of commercial arts, are exposed to corruption, by their admitting wealth, unsupported by personal elevation and virtue, as the great foundation of distinction, and by having their attention turned on the side of interest, as the road to consideration and honour.' (Ferguson 1995: 241).

Ambition is ubiquitous and so will drive people to acquire what they regard as material improvement. We cannot blame the division of labour and the pursuit of material comfort for moral corruption: it can occur at any level of wealth. As he puts it: 'the vices of men are not proportioned to their fortunes' (Ferguson 1995: 236). Again, Ferguson is absolutely clear in this: 'we may find him become effeminate, mercenary, and sensual; not because pleasures and profits are becoming more alluring, but because he has fewer calls to attend to other objects...' (Ferguson 1995: 237). This is the heart of Ferguson's account. The problem is not self-regarding

behaviour or ambition: it is a failure of that ambition to be excited by both material and moral objects.

Ferguson's concern is not that commerce will lead to moral corruption as self-interest overwhelms benevolence. Instead the worry is that the peace of a commercial society will rob us of an arena in which to exercise our moral ambition. The absence of 'public alarms' (Ferguson 1995: 244) creates the space for a corruption of the objects of ambition.<sup>13</sup> A society where the people face no calls to act in the public interest is a society that will create the space for the corruption of the object of ambition. It is not the material goods, or the division of labour that is the problem, but rather a political situation bereft of opportunities to distinguish ourselves through public action.

The problem arises when this self-interest extends into selfishness, when it encompasses all of our attention and becomes 'the sole object of human care' (Ferguson 1994: 99). Excessive focus on interest can distract us from society, but even then human nature is so deeply social that we are always associating with others and 'Even while the head is occupied with projects of interest, the heart is often seduced into friendship.' (Ferguson 1995: 40). So self-preservation can become selfishness through a 'mistake in the choice of our object' (Ferguson 1995: 54). There is nothing particularly novel or interesting about Ferguson's view on this. He is following Hutcheson's response to Mandeville and is treading the same turf as Hume and Smith's attempts to reconfigure the virtues to downplay the dangers of commercial activity. He is also following Hutcheson when he comes to discuss the nature of the law of society as manifested in the principle of benevolence.

One way we can understand what Ferguson is arguing here is by looking at his account of the 'voluptuary'. Ferguson distinguishes between the mistaken life of the 'innocent' voluptuary and the life of one 'deeply infected with malice, envy, and pride' (Ferguson 1973 Vol. 2: 77), but in doing so he gives perhaps the clearest statement that the problem is not wealth but a lack of virtue: 'We err, in deriving the corruptions, which are imputed to great cities and courts, from the love of pleasure, and from the profusion of wealth, with which the love of pleasure is gratified.' (Ferguson 1973 Vol. 2: 77). The problem is not wealth, or even the love of wealth, it is when these are not accompanied by other objects of ambition. As he puts it:

The wants of nature are easily supplied, and the gratification of uncorrupted appetite is fully consistent with the higher and better pursuits of human life: But the voluptuary in acquiring habits inconsistent with these pursuits, is debauched by his imagination rather than by the force of his appetite, or by the solicitations of Sense (Ferguson 1973 Vol. 2: 342).

The problem is a society where wealth entirely replaces honour and public duty as the means of advancing ambition. The argument in the *Essay* is not based on the idea that the self-interest of commercial actors leads to moral corruption, instead his argument is that it is not commerce, but politics that is the problem. It is a political failure created by the absence of calls to public attention that is the root of Ferguson's concern here. That is to say that it is not the advent of commerce, but the absence

---

<sup>13</sup>A point made by Brandon Turner (2012) who notes Ferguson's interest in the non-political yet public realm that we have come to call Civil Society.

of politics that is the root issue. There is no reason in principle why a commercial society need not also be a politically engaged society.<sup>14</sup>

The incentive to engage in politics is also ambition. It is personal aggrandisement, but aggrandisement earned from service to the public rather than the display of wealth. I achieve the renown I seek by acting in the interest of the community rather than by displaying my wealth: the vices of the rich are not the result of excess wealth, but of having nothing to do, it is inactivity rather than wealth that is the issue for Ferguson. Ambition motivates the statesman and the warrior, but it is ambition for themselves within the wider group identity.

Ferguson goes further than this. It is not enough to argue that the absence of calls to public action allows the space for moral corruption. This also has to be seen in the context of the system of Ferguson's use of the moralised language of the civic tradition in connection with a theory of rank.

The subjects of property, considered with a view to subsistence or even to enjoyment, have little effect in corrupting mankind, or in awakening the spirit of competition and of jealousy; but considered with a view to distinction and honour, where fortune constitutes rank, they excite the most vehement passions, and absorb all the sentiments of the human soul: they reconcile avarice and meanness with ambition and vanity; and lead men through the practice of sordid and mercenary acts to the possession of a supposed elevation and dignity (Ferguson 1995: 154).

In a sense then Ferguson is admitting the reality of Mandeville's view from the *Female Tatler*, that we engage in public spirited behaviour from ambition, it is just that Ferguson can apply that term in a more sophisticated and less pejorative sense. The primary motive to public service here is quite clear, it is not benevolence, but rather ambition. The danger is not that it is present, but that it can be reconciled with avarice and meanness. In a society where there are multiple objects of distinction this is not a problem. A society where one may gain acclaim through wealth, through military service, and through political action is one where different routes to prominence are open. The balance of different routes to status that serve as the objects of ambition indicates a successfully balanced society.

In a sense we have come full circle: Ferguson seems to be making standard republican arguments drawn from Machiavelli about the need for active citizens and stable institutions. But even here, as Max Skjönsberg has recently pointed out, all is not quite what it seems. Ferguson's account is based on the absolute centrality of a variety of routes to public acclaim. Rank from birth, from wealth, from service and from leadership are all required to create a stable political order. Each order must do its bit. And it is here where the republican reading of Ferguson becomes strained. As Lisa Hill (2006: 27) has pointed out, Ferguson's political conservatism appears to sit uneasily with his purported republicanism. Ferguson was no radical, and in fact the idea of a society where the only route to esteem and rank is political activity is just as anathema to him as one where wealth alone, or military service alone, are the sole routes to distinction. Human beings are ambitious and a stable society needs to provide them with a range of objects for that ambition.

---

<sup>14</sup>A point made at some length in my (2008) paper on Ferguson's conception of the active life.

Though we should add a caveat here. Ferguson was worried by the idea that someone who committed themselves to a political ideal that they saw as in the public interest might become too politically active and develop revolutionary fervour. ‘The Zealot for liberty has run into the wildest disorders, and adventures, and under pretence of promoting it, have found their way to the most violent and pernicious usurpations.’ (Ferguson 1973 Vol. 2: 457).

## 5 Conclusion

My aim in this paper has been to suggest that the reading of Ferguson as a figure who, when compared to Hume and Smith, was more sceptical of commerce and who mined the Republican and Civic Humanist traditions to make his case, is a partial reading of his view. Moreover, it is a reading that becomes obviously partial when one moves beyond the *Essay* and focuses on his overall project. I have tried to understand Ferguson’s views on self-interest within his wider system of moral science and in doing so hope to have shown that his analysis is more subtle than we might think if we read him chiefly as a republican foil for Hume and Smith or as a precursor of Marxian analysis.

Comparing Ferguson to Hume and Smith on commerce, with a focus on the *Essay*, has created a misleading impression of the place of self-interest in his thought. Just like his friends, Ferguson believed that both self-interest and benevolence were part of human behaviour and, as a result, an accurate and empirically grounded moral philosophy, had to deal with this fact and explain the proper relationship between them. In his development of the idea of moral ambition, Ferguson takes a different path from Hume and Smith, but the broad enterprise and the end position he reaches on the value of commercial society is much the same as the other two. His concept of self-interest is moralised, but it is not purely negative. Appropriate self-regarding behaviour in the form of moral and material ambition in their proper balance lies at the heart of Ferguson’s account of a commercial society.

## References

- Allan, David. 2006. *Adam Ferguson*. Aberdeen: AHRC Centre for Scottish and Irish Studies.
- Arbo, Matthew B. 2011. Adam Ferguson’s Sermon in the Ersh Language: A word from 2 Samuel on martial responsibility and political order. *Political Theology* 12 (6): 894–908.
- Benton, Ted. 1990. Adam Ferguson and the enterprise culture. In *The enlightenment and its shadows*, ed. P. Hulme and L. Jordanova, 101–20. London: Routledge.
- Berry, Christopher J. 1994. *The idea of luxury: A conceptual and historical investigation*. Cambridge: Cambridge University Press.
- Berry, Christopher J. 2009. But art itself is natural to man’: Ferguson and the principle of simultaneity. In *Adam Ferguson: Philosophy*, ed. Eugene Heath and Vincenzo Merolle, 143–53. Pickering and Chatto: Politics and Society, London.

- Berry, Christopher J. 2013. *The Idea of Commercial Society in the Scottish Enlightenment*. Edinburgh: Edinburgh University Press.
- Bottomore, Tom, and Robert Nisbet. 1978. *A history of sociological analysis*. London: Heinemann.
- Brewer, John D. 1986. Adam Ferguson and the theme of exploitation. *The British Journal of Sociology* 37(4): 461–78.
- Brewer, John D. 1989. Conjectural history, sociology and social change in eighteenth century Scotland: Adam Ferguson and the division of labour. In *The making of Scotland: Nations*, ed. D. McCrone, S. Kendrick, and P. Straw. Culture and Social Change, Edinburgh: Edinburgh University Press, pp. XXX.
- Brewer, John D. 2007. Putting Adam Ferguson in his place. *The British Journal of Sociology* 58(1): 105–22.
- Ferguson, Adam. 1756. *Reflections previous to the establishment of a Militia*, London R. and J. Dodsley.
- Ferguson, Adam. 1766. *Analysis of pneumatics and moral philosophy*, Edinburgh: A. Kincaid and J. Bell.
- Ferguson, Adam. 1856. *The history of the progress and termination of the Roman Republic*, New York: J.C. Derby. University of Michigan: Historical Reprint Series.
- Ferguson, Adam. 1973 [1792]. *Principles of moral and political science*, 2 vols. New York: AMS Press.
- Ferguson, Adam. 1994 [1769]. *Institutes of moral philosophy*. London: Routledge/Thoemmes Press.
- Ferguson, Adam. 1995 [1767]. *An essay on the history of civil society*, ed. Fania Oz-Salzberger. Cambridge: Cambridge University Press.
- Ferguson, Adam. 1995. *The correspondence of Adam Ferguson*, 2 vols. ed. Vincenzo Merolle. London: William Pickering.
- Ferguson, Adam. 2006. *The manuscripts of Adam Ferguson*, ed. Vincenzo Merolle, London: Pickering and Chatto.
- Forbes, Duncan. 1982. Natural law and the Scottish enlightenment. In *The origins and nature of the Scottish enlightenment*, ed. R.H. Campbell and A. Skinner, 186–204. John Donald: Edinburgh.
- Goldsmith, M.M. 1985. *Private vices, public benefits: Bernard Mandeville's social and political thought*. Cambridge: Cambridge University Press.
- Goldsmith, M.M. 1988. Regulating anew the moral and political sentiments of mankind: Bernard Mandeville and the Scottish Enlightenment. *Journal of the History of Ideas* 49 (4): 587–606.
- Graham, Gordon. 2013. Adam Ferguson as a Moral Philosopher. *Philosophy* 88(4): 511–25.
- Hamowy, Ronald. 1968. Adam Smith, Adam Ferguson, and the Division of Labour. *Economica* XXXV: 249–59.
- Hill, Jack A. 2013. Marx's reading of Adam Ferguson and the idea of progress. *Journal of Scottish Philosophy* 11(2): 167–90.
- Hill, Jack A. 2017. *Adam Ferguson and ethical integrity: The man and his prescriptions for the moral life*. Lanham, MD: Lexington Books.
- Hill, Lisa. 2001. Eighteenth-century anticipations of the sociology of conflict: The case of Adam Ferguson. *Journal of the History of Ideas* 62(2): 281–299.
- Hill, Lisa. 2006. *The passionate society: The social political and moral thought of Adam Ferguson*. Dordrecht: Springer.
- Hill, Lisa. 2007. Adam Smith, Adam Ferguson and Karl Marx on the Division of Labour. *Journal of Classical Sociology* 7(3): 339–66.
- Hume, David. 1985. *Essays moral, political, and literary*, ed. Eugene F. Miller. Indianapolis: Liberty Fund.
- Hundert, E.J. 1994. *The enlightenment's fable: Bernard Mandeville and the discovery of society*. Cambridge: Cambridge University Press.
- Hutcheson, Francis. 1750. *Reflections upon laughter, and remarks on the fable of the bees*. Glasgow: R. Urie.
- Hutcheson, Francis. 2002 [1728]. *An essay on the nature and conduct of the passions and affections, with illustrations on the moral sense*, ed. Aaron Garrett. Indianapolis: Liberty Fund.



- Hutcheson, Francis. 2004 [1726]. *An inquiry into the original of our ideas of beauty and virtue*, ed. Wolfgang Leidhold. Indianapolis: Liberty Fund.
- Kalyvas, Andreas and Katznelson, Ira. 1998. Adam Ferguson returns: Liberalism through a glass, darkly. *Political Theory* 26(2): 173–97.
- Kettler, David. 2005. *Adam Ferguson: His social and political thought*. New Brunswick and London: Transaction Press.
- MacRae, Donald G. 1969. Adam Ferguson. In *The founding fathers of social science*, ed. T. Raison, 17–26. Harmondsworth: Penguin.
- Mandeville, Bernard. 1971 [1732]. An enquiry into the origin of honour and the usefulness of christianity in War. 2nd ed., ed. M.M. Goldsmith, London: Frank Cass.
- Mandeville, Bernard. 1988. *The fable of the bees or private vices, publick benefits*, ed. F.B. Kaye. Indianapolis: Liberty Fund.
- Mandeville, Bernard. 1999. [1709–10]. *By a society of ladies: Essays in the female tatler*, ed. M. M. Goldsmith. Bristol: Thoemmes Press.
- McDaniel, Iain. 2013. *Adam Ferguson in the Scottish enlightenment: The roman past and Europe's future*. Cambridge Mass: Harvard University Press.
- McDowell, Gary L. 1983. Commerce, virtue, and politics: Adam Ferguson's constitutionalism. *The Review of Politics* XLV: 536–552.
- Meek, Ronald. 1967. The Scottish contribution to Marxist sociology. In *Economics and ideology and other essays*, ed. R.L. Meek, 34–50. London: Chapman and Hall.
- Merolle, Vincenzo. 2009. Hume as a critic of Ferguson's essay. In *Adam Ferguson: philosophy, politics and society*, ed. Eugene Heath and Vincenzo Merolle, 73–87. Pickering and Chatto: London.
- Mizuta, Hiroshi. 1981. Two Adams in the Scottish enlightenment: Adam Smith and Adam Ferguson on progress. *Studies in Voltaire* 191: 182–99.
- Monro, Hector. 1975. *The ambivalence of Bernard Mandeville*. Oxford: Clarendon.
- Pocock, J.G.A. 1975. *The Machiavellian moment: Florentine political thought and the Atlantic Republican tradition*. Princeton NJ: Princeton University Press.
- Pocock, J.G.A. 1999. *Barbarism and religion: Narratives of civil government*, vol. II. Cambridge: Cambridge University Press.
- Raynor, David. 2009. Why did David Hume dislike Adam Ferguson's an essay on the history of civil society. In *Adam Ferguson: philosophy, politics and society*, ed. Eugene Heath and Vincenzo Merolle, 45–72. Pickering and Chatto: London.
- Reeder, John. 1997. *On moral sentiments: Contemporary responses to Adam Smith*. Bristol: Thoemmes Press.
- Robertson, John. 1985. *The Scottish enlightenment and the Militia issues*. Edinburgh: John Donald.
- Rutherford, Donald. 2012. *In the shadow of Adam Smith: founders of Scottish economics 1700–1900*. London: Palgrave Macmillan.
- Sebastiani, Silvia. 2011. National character and race: A Scottish enlightenment debate. In *Character, self, and sociability in the Scottish enlightenment*, ed. T. Ahnert and S. Manning, 187–205. London: Palgrave MacMillan.
- Sebastiani, Silvia. 2013. *The Scottish enlightenment: Race, gender, and the limits of progress*. London: Palgrave MacMillan.
- Sher, Richard B. 1985. *Church and University in the Scottish enlightenment: The moderate literati of Edinburgh*. Edinburgh: Edinburgh University Press.
- Sher, Richard B. 1989. Adam Ferguson, Adam Smith and the problem of National Defence. *Journal of Modern History* 61: 240–268.
- Sher, Richard B. 1994. From Troglodytes to Americans: Montesquieu and the Scottish enlightenment on liberty, virtue, and commerce. In *Republicanism, liberty and commercial society*, ed. D. Wooton, 368–402. Stanford, CA: Stanford University Press.
- Skinner, Andrew. 1982. A Scottish contribution to Marxist sociology? In *Classical and Marxian sociology*, ed. I. Bradley and M.C. Howard, 95–121. London: Macmillan.

- Skjöönsberg, Max. 2017. Adam Ferguson on Partisanship, party conflict, and popular participation. *Modern Intellectual History*: 1–28. <https://doi.org/10.1017/s14792244317000099>.
- Smith, Adam. 1976 [1759]. *The theory of moral sentiments*, ed. D.D. Raphael and A.L. Macfie. Oxford: Oxford University Press.
- Smith, Adam. 1980 [1795]. *Essays on philosophical subjects*, ed. W.P.D. Wightman. Oxford: Oxford University Press.
- Smith, Adam. 1987. *The correspondence of Adam Smith*, ed. E.C. Mossner and I.S. Ross. Oxford: Oxford University Press.
- Smith, Craig. 2008. Ferguson and the Active Genius of Mankind. In *Adam Ferguson: History, progress and human nature*, ed. Eugene Heath and Vincenzo Merolle, 157–70. London: Pickering and Chatto.
- Smith, Craig. 2018. *Adam Ferguson and the idea of a civil society: Moral science in the Scottish enlightenment*. Edinburgh: Edinburgh University Press.
- Swingewood, Alan. 1970. Origins of sociology: The case of the Scottish enlightenment. *British Journal of Sociology* 21: 164–180.
- Turner, Brandon P. 2012. Adam Ferguson on “Action” and the possibility of non-political participation. *Polity* 44: 212–33.
- Winch, Donald. 1978. *Adam Smith’s politics: An essay in historiographic revision*. Cambridge: Cambridge University Press.

# Adam Smith on Self-interest



Shinji Nohara

**Abstract** Adam Smith extolled various virtues including benevolence in his *The Theory of Moral Sentiments*, although he focused primarily on self-interest in *The Wealth of Nations*. Whereas human nature as discussed in both works is compatible theoretically, Smith, in his narrative, described different aspects of human nature in each work. However, the difference in Smith's descriptions of human action in *The Theory of Moral Sentiments* and *The Wealth of Nations* is not actually a conflict between morality and economy. Economic behavior is based on a jurisprudential framework in which one can protect one's property. For Smith, protecting one's property is the basis of natural rights and is the foundation of social relationships in political economy; although individuals prefer voluntary donation and hospitality, these virtues are not the basis of social existence, which is necessary for protecting one's property. Based on the protection of one's property, one can pursue self-interest. As a result, Smith's *Lectures on Jurisprudence* is the key to understanding the connection between *The Theory of Moral Sentiments* and *The Wealth of Nations*. Indeed, Smith regarded *The Wealth of Nations* as a part of his jurisprudence. His jurisprudence was based on *The Theory of Moral Sentiments*. If we understand the relationship among the three works, we can more properly understand the different human descriptions of human nature in the three works more properly. Through this understanding, we can perceive the foundations of economic behavior.

**Keywords** Adam Smith · Jurisprudence · Self-interest · The Wealth of Nations · The Theory of Moral Sentiments

## 1 Introduction

This chapter examines why Adam Smith (1723–90) focused on self-interest in describing human nature in *An Inquiry into the Nature and Causes of the Wealth of Nations* (hereafter *WN*) in 1776, although in *The Theory of Moral Sentiments*

---

S. Nohara (✉)  
University of Tokyo, Tokyo, Japan  
e-mail: [nohara@g.ecc.u-tokyo.ac.jp](mailto:nohara@g.ecc.u-tokyo.ac.jp)

(hereafter *TMS*) in 1759, he explained various kinds of human nature, many of which were not based on self-interest. Whereas both works are theoretically coherent, the different descriptions of human nature in both works remain, as many scholars on Smith have argued. As I argue in this chapter, to answer this question, it is important to pay attention to Smith's system of jurisprudence. Indeed, *WN* is part of his jurisprudence, which is based on *TMS*. This jurisprudential framework of human nature influenced the human nature in *WN*.

This point is important in considering human nature in economic behavior. Self-interest is seen as the foundation of human nature in economics. Smith, normally regarded as the founder of classical economics, and known as one of the first to examine the market mechanism, is regarded as characterizing human nature as self-interested. As Jacob Viner wrote, Smith supposedly thought that, "in the market...man behaves in response to calculating, rational self-interest" (Viner 1960, 60). Self-interest or economic rationality is the foundation of economic behavior in the market, and especially today, many people think of those in the market as self-interested. Further, economists presuppose that human nature in terms of the market mechanism is based on economic rationality, in which self-interested people rationally calculate their own interests.

In spite of the fact that Smith is viewed as the founder of the theory of the market mechanism, many scholars do not see him as assuming economic rationality as the sole motive. Certainly, in *WN*, especially in Book I, Chap. II, Smith stated, "It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest" (*WN*, I. ii. 2). Nevertheless, Smith was also the author of *TMS*, which does not characterize self-interest as the sole or main motivator of human beings. This problem was known as *Das Adam Smith Problem* in nineteenth-century Germany: the problem of why the author of *WN*, who emphasized self-interest, wrote *TMS*, where he was seen as advocating altruistic human nature (see Montes 2004, Chap. II). Today, many scholars do not think of the works as inconsistent in describing human nature. As Raphael and Macfie (1982) pointed out, *WN* and *TMS* are not incompatible because sympathy in *TMS* is not the same as altruism and is also the foundation of self-interest in *WN*.

Nonetheless, recent scholars have found discrepancies between *TMS* and *WN*. For example, Brown contrasted *TMS* as a dialogical text with *WN* as a monological text (Brown 1994, Chap. II). Pack compared the ahistorical approach of *TMS* with the historical approach of *WN* (Pack 1997). Otteson wondered why *WN* focused on self-interest as the sole motivator of human nature, whereas Smith in *TMS* mentioned various human motives (Otteson 2002, Chap. IV). Nieli believed that human nature in *WN* could be based on self-interest because the human relationships in *WN* were not intimate or based on friendship, unlike in *TMS* (Nieli 1986). Griswold discussed how human nature in *WN* is related to moral theory (Griswold 1999, 263–64; also see Young 1997, 173). In another perspective, Montes argued how sympathy leads not only to moral judgment but also to behavior and is comprehensible as a motive (Montes 2002, Chap. II). It remains to be clarified why *WN*, in its description of human nature, centered in self-interest, and *TMS* did not; even though *WN* and

*TMS* are theoretically compatible, they comprehended human nature with different emphases.

To resolve this issue, it is important to consider *WN* as a part of Smith's jurisprudence, which was based on his description of human nature and morals in *TMS*. Thus, human nature in *WN* is directly linked with *TMS* through Smith's jurisprudence. His work on jurisprudence was not published, but lecture notes transcribed by his students remain as two sets of *Lectures on Jurisprudence* (hereafter *LJ*; *LJ(A)* is "Report of 1762–1763," and *LJ(B)* is "Report dated 1776"). Indeed, Smith wrote *WN* as a part of his natural jurisprudence. He discussed, "the general principles of law and government, and of the different revolutions which they had undergone in the different ages and periods of society; not only in what concerns justice, but in what concerns police, revenue, and arms, and whatever else is the object of law" (*TMS*, Advertisement). Natural jurisprudence "ought to run through and be the foundation of the laws of all nations" (*TMS*, VII. iv. 37). By contrast, "systems of positive law...can never be regarded as accurate systems of the rules of natural justice" (*TMS*, VII. iv. 36). For Smith, natural justice was the basis of his natural jurisprudence, including *WN* (see Young 1997).

Smith's jurisprudence has its own context. Although natural jurisprudence before Smith often referred to economic subjects, it did not contain much economic theory. Therefore, Smith tried to transform natural jurisprudence. In the latter half of the eighteenth century, some French thinkers after Jean-Jacques Rousseau also tried to transform natural jurisprudence into social science (Sonenscher 2009).

The fact that a natural jurisprudential framework influenced Smith's thought, including *WN*, is important in understanding the treatment of self-interest in *WN*. Certainly, the self-interest in *WN* was depicted in *TMS*. Nevertheless, in Smith's system of moral philosophy, *WN* rests on his jurisprudential system. Thus, Smith's jurisprudential framework influenced his depiction of human nature in *WN*.

It is important to note that natural rights are the basis of natural jurisprudence. Whereas *TMS* is the foundation of the explanation of human nature in Smith's overall system, he proposed in *WN* that natural rights define the basic framework for depicting human nature in *WN*. Linking political economy to natural jurisprudence was what Smith intended to do. While many secondary sources have suggested that Smith characterized human nature consistently in *WN* and *TMS* (Raphael and Macfie 1982; Otteson 2002; Montes 2004; Raphael 2007), these studies have not examined the jurisprudential impact in *WN*. Certainly, some paid attention to Smith's works other than *WN* and *TMS*, but they did not do so adequately with respect to the depiction of human nature in his jurisprudence (Brown 1994; Göçmen 2007). Smith wrote largely on economic theory in the 'police' section on natural jurisprudence. The basis of natural jurisprudence for Smith is justice. In this sense, *WN* is the extension and transformation of his theory of justice in natural jurisprudence. This is important in elucidating human nature and self-interest in *WN*.

Here, I would like to discuss why the term "self-interest" is used in this essay. Certainly, Smith did not define self-interest, self-love, and selfishness. For instance, Pierre Force argued that "self-interest," rarely used in *WN*, means vanity, whereas "self-love" means self-preservation (Force 2003, 42–47). Additionally, Griswold

argued that “selfishness” means self-centeredness (Griswold 1999, 81). Whereas selfishness as self-centeredness focuses more on one’s inability to know others directly, and self-love focuses on self-preservation, self-interest focuses more on one’s regard for one’s property than self-centeredness or self-love.

Overall, understanding the background of self-interest as referenced in Smith contributes to deepening our recognition of the foundations of the economic motives in *WN*. Although Smith described various human motives in *TMS*, he emphasized a limited human motive such as self-interest in *WN* because of his jurisprudential basis of human nature in economic behavior.

This chapter intends to analyze that jurisprudential basis in *WN*. It is a part of Smith’s natural jurisprudence and view on natural rights. Natural rights are linked to self-interest, which is the basis of human depiction in *WN*. The next section examines the context of arguing self-interest in view of natural rights, and the third section investigates how Smith treated natural rights in his natural jurisprudence; the fourth section addresses self-interest in economic behavior.

## 2 Natural Rights and Self-interest Before Smith

Knowledge of the jurisprudential framework of humanity is a prerequisite for understanding the discussion of human nature in *WN*. This section considers how the treatment of natural rights was connected to self-interest before Smith.

First, I would like to examine the context of Smith’s depiction of natural rights. In the early modern period, there was a divergence on the theoretical foundation of natural jurisprudence; while Hugo Grotius and Thomas Hobbes regarded natural rights as the basis of natural jurisprudence, others such as Samuel Pufendorf denied it and instead saw duty as the basis of natural jurisprudence (Tuck 1979).

For Grotius, “natural right is the dictate [sic] of right reason, shewing the moral turpitude, or moral necessity, of any act from its agreement or disagreement with a rational nature, and consequently that such an act is either forbidden or commanded by God, the author of nature” (Grotius 1901, 21). Natural rights can be discovered through reason. The reasonableness of natural rights is sustained by God, the creator of the world. Grotius’s natural rights are ultimately authenticated by God. This point is similar to the advocates of natural rights in the Middle Ages such as Jean Gerson who argued that self-preservation was the basis of natural law and rights, which were ultimately authenticated by God (Tierney 2001: 322).

Natural rights are the foundation of the right to protect property, which is the main object of legal protection. As Grotius wrote,

Natural right relates not only those things that exist independent of the human will, but to many things, which necessarily follow the exercise of that will. Thus property, as not in use, was at first a creature of the human will. But, after it was established, one man was prohibited by the law of nature from seizing the property of another against his will (Grotius 1901, 21).

Grotius (1901, 86–88) also explained the origin of private property: “All things ... formed a common stock for all mankind, as the inheritors of one general patrimony.” In the primitive period, people needed no private property. They were exempted from human evils such as appetites. However, as time went by, “the greatest breach in the harmony of men was made by ambition.” First, during the time of pasture, movables such as herds came to be seen as private property, and after the introduction of agriculture, the land became private property (Grotius 1901, 86–88). As a result, “from hence it happened, that every man seized to his own use or consumption whatever he met with; a general exercise of a right, which supplied the place of private property. So that to deprive any one of what he had thus seized, became an act of injustice” (Grotius 1901, 86). Thus, “property...must have been established either by express agreement, as by division, or by tacit consent, as by occupancy” (Grotius 1901, 89). As this suggests, the protection of private property means the protection of one’s proper things: the concept of property is based on self-interest.

Hobbes also regarded natural rights as the foundation of social maintenance. He based natural rights on his social contract theory; He said,

The RIGHT of NATURE, which Writers commonly called Jus Naturale, is the liberty each man hath, to use his own power, as he will himself, for the preservation of his own Nature; that is to say, of his own LIFE; and consequently, of doing any thing, which in his own Judgement, and Reason, hee [sic] shall conceive to be the aptest means thereunto. (Hobbes 1991, 91)

He explained what a right was: “to lay down a mans Right to any thing, is to devert himselfe [sic] of the Liberty, of hindering another of the benefits of his own Right to the same” (Hobbes 1991, 92). Natural rights are based on protecting one’s own things, and the protection presupposes self-interest defined as pursuing the possession of things.

Francis Hutcheson criticized Hobbes and other students of natural law who saw human nature as self-interested. Hutcheson wrote, “the greatest part of our latter moralists” like Hobbes “establish it as undeniable...since all Laws operate only by Sanctions of Rewards, or Punishments, which determine us to Obedience by Motives of Self-Interest.” They supposed, “that it is thus that Laws do constitute some Actions mediately Good, or Advantageous, and others the same way Disadvantageous.” They say indeed, “That a benevolent Legislator constitutes no Actions Advantageous to the Agent by Law” (Hutcheson 2004, 87). Hutcheson thus criticized Hobbes’s notion of a self-interested human nature because a legislator cannot not make people obey laws only through his benevolence. For Hobbes, punishment and reward were necessary for maintaining laws. Hutcheson criticized this view of laws because it presupposed a self-interested human nature. For Hobbes, without self-interested considerations of punishment and reward, people were not supposed to obey laws. Hutcheson opposed to this view of human nature (Hutcheson 2004, 93–94).

Unlike Hobbes, Hutcheson insisted that human beings have an innate sense of right and wrong. He wrote,

Since it was God our Creator who implanted this sense of right and wrong in our souls, and gave us these powers of reason...observing our own constitution, and that of persons and other things around us, discovers what conduct tends either to the common prosperity of all, or that of individuals, and what has a contrary tendency...all these precepts or practical dictates of right reason are plainly so many laws, enacted, rectified by penalties, and promulgated by God in the very constitution of nature (Hutcheson 1747, 112).

Also for Hutcheson, unlike Hobbes, natural rights were not based on self-interest. He wrote, “since it is manifestly necessary to the common interest of all that large numbers of men should be joined together in amicable societies...we therefor [sic] deem that each man has a right to procure and obtain for himself or his friends such advantages and enjoyments” (Hutcheson 1747, 118). As a result, “no private right can hold against the general interest of all. For a regard to the most extensive advantage of the whole system ought to controll [sic] and limit all the rights of individuals or of particular societies” (Hutcheson 1747, 120).

For Grotius and Hobbes, natural rights are linked directly with self-interest, whereas Hutcheson opposed the self-interested foundation of natural rights. Then, Hutcheson saw natural rights as limited.

### 3 Natural Rights and Self-interest in Smith

Smith faced a problem, that is, how to reconcile a fully-fledged respect for natural rights with a concept of human nature not based on self-interest. To solve the problem, Smith explained how natural rights develop naturally from sympathy. Through sympathy, people form morals and, through morals, form a sense of natural rights.

Smith described that natural rights are derived from sympathy, by elucidating how, through sympathy, people form moral sentiments. Sympathy is a feeling of fellowship towards others. However, because one cannot know another person’s inner sentiments, one imagines and sympathizes with the other person’s sentiments through an imagined change in positions, which makes it possible to think about the other person’s behavior as if one is the other person. Smith wrote, “as we have no immediate experience of what other men feel, we can for no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation” (*TMS*, I. i. 1. 2). Similarly, one also wants the sympathy of others. He mentioned, “the sympathy, which my friends express with joy, might, indeed, give me pleasure by enlivening that joy” (*TMS*, I. i. 2. 2). Then, “we are still more anxious to communicate to our friends our disagreeable than our agreeable passions, that we derive still more satisfaction from their sympathy with the former than from that with the latter, and that we are still more shocked by the want of it” (*TMS*, I. i. 2. 3). He also said, “love is an agreeable; resentment, a disagreeable passion; and accordingly we are not half so anxious that our friends should adopt our friendships, as that they should enter into our resentments,” and in particular “we can forgive them [friends] though they seem to be little affected with the favours which we may have received,



but lose all patience if they seem indifferent about the injuries which may have been done to us” (*TMS*, I. i. 2. 5). Thus, people sympathize with the resentment of a victim when an injury is reported.

Resentment is the origin of justice. When an agent A injures person B, the spectator sympathizes with B and feels resentment towards A and therefore thinks that A should be punished. This is the origin of the feeling of justice (*TMS*, II. i. 2. 5). This sense of justice is based on moral sentiments: “there can be no proper motive for hurting our neighbour, there can be no incitement to do evil to another, which mankind will go along with, except just indignation for evil which that other has done to us” (*TMS*, II. ii. 2. 1). People require justice because of their resentment.

The most sacred laws of justice... those whose violation seems to call loudest for vengeance and punishment, are the laws which guard the life and person of our neighbour; the next are those which guard his property and possessions; and last of all come those which guard what are called his personal rights, or what is due to him from the promises of others. (*TMS*, II. ii. 2. 2)

Thus, justice depends on the protection of one’s proper things, or natural rights. Injury produces resentment. In resentment, people require justice, or the protection of one’s own natural rights. Thus, Smith analyzed how people form natural rights through sympathy.

Natural rights for Smith were not based on Hobbes’ social contract, in which people agree on what their natural rights are, nor on Hutcheson’s view that people form a sense of natural rights due to their sense of the general interests of society. These approaches rest on the condition that people directly recognize natural rights. For Smith, people do not explicitly recognize natural rights, but they understand natural rights through the sympathetic process of the formation of justice by resentment.

Smith made clear the relationship between moral sentiments and natural rights in his *LJ*. Indeed, injury, according to Smith, begins with natural rights. First, in terms of injury concerning one’s person, Smith wrote, “a man may be injured in his person two ways also, either 1st, by killing, wounding, or maiming him, or any way hurting his body, or secondly by restraining his liberty... These rights correspond to what Pufendorff [sic] call < s > naturall [sic] rights” (*LJ(A)*, i. 12). Especially, “those which affect his body may all be reduced to the first class of those respecting his person” (*LJ(A)*, i. 13). Second, about injury to one’s reputation, he stated that “a man is injured in his reputation when one endeavours to bring his character below what is the common standard amongst men” (*LJ(A)*, i. 13). Third, a man may be injured on his estate (*LJ(A)*, i. 16). As a result, natural rights are invoked when an injury occurs.

Then, Smith categorized three types of natural rights. He wrote, “we shall consider in the first place those rights that belong to a man as a man, as they are generally most simple and easily understood, and generally can be considered without respect to any other condition.” In concrete terms, “a man merely as a man may be injured in three respects, either 1st, in his person; or 2dly, in his reputation; or 3dly, in his estate” (*LJ(A)*, i. 11–12).

When Smith explained the formation of natural rights, he saw them as being formed *a posteriori*; that is, an injury led to resentment through which justice and

one's proper things were made clear. As a result, natural rights were formed in the natural process of human interaction. One's proper things existed not as antecedent to human relationships but following from them. As he said, "the concern which we take in the fortune and happiness of individuals does not, in common cases, arise from that which we take in the fortune and happiness of society...our regard for the multitude is compounded and made up of the particular regards which we feel for the different individuals of which it is composed" (*TMS*: II. ii. 3. 10).

Therefore, justice is unique among the virtues. Unlike other virtues, justice may be obligatory (*TMS*, II. ii. 1. 5). The *a posteriori* formation of justice does not mean that it was ambiguous: "the rules of justice are accurate in the highest degree, and admit of no exceptions or modifications" (*TMS*, III. 6. 10). By contrast, people feel benevolence, compassion, and other warm virtues towards their company rather than towards strangers (*TMS*, VI. ii. 1. 17).

Justice should be observed strictly, so it can serve as the basis of protecting natural rights. Natural rights can then be divided into perfect and imperfect rights. Because, unlike justice, people know that someone might not express or exercise beneficence and compassion, these virtues are seen as imperfect rights: "what they call imperfect rights are those which correspond to those duties which ought to be performed to us by others but which we have no title to compel them to perform" (*LJ(A)*, i. 14). On the other hand, justice can be compulsory and is connected to perfect rights: "perfect rights are those which we have a title to demand and if refused to compel an other to perform" (*LJ(A)*, i. 14). Unlike beneficence or other virtues of imperfect rights, perfect rights can require others to protect one's proper things.

Justice depends on the protection of one's proper things and thus presupposes a sense of propriety. Basically, for Smith, virtues including justice do not exist *a priori* but are formed by a sense of propriety produced by the interaction of people. Originally, the term "propriety" was derived from the Latin term "proprius," which meant belonging only to one's own. Property was a doublet, and meant one's proper. In the seventeenth century, although propriety was a doublet for property, the former had a broader meaning than the latter. In the eighteenth century, the concept of propriety had moral implications different from its property implications (Montes 2004: 98–101). The concept of propriety formed the basis of the standard for judging morality but was not moral itself.

This formation of the sense of propriety was similar to that of property since people foster the sense of property in human relationships, especially in relation to governments as the executors of justice in *LJ*. Smith said, "the first and chief design of all civil [sic] government, is, as I observed, to preserve justice amongst the members of the state and prevent all incroachments on the individuals in it, from others of the same society" (*LJ(A)*, i. 9). This "as I observed" meant his lecture on morality, which resulted in *TMS*. As a result, property in *LJ* was the logical consequence of his consideration of justice in *TMS*. In *LJ*, the rights of such one's own lives and bodies were universal across all ages and societies. However, "the only case where the origin of naturall [sic] rights is not altogether plain, is in that of property" (*LJ(A)*, i. 25).

People in each society have each different senses of their own things as property. When Smith distinguished four developmental stages of society, he described each different sense of property. In the first hunters' society, there was almost no notion of property. In the next pastoral society, when herds were kept as the way of subsistence, herds became one's property, and governments were formed to protect them. In the third, agricultural society, where land was cultivated as the source of subsistence, land became the object of one's property. In the final commercial society, there arose a great number of commodities and trades, producing a wide variety of types of property (*LJ(A)*, i. 26–32). As a result, he established the four stages of society and each stage differs in its forms of property. Laws also differ according to each of these stages. *LJ* related laws to their social backgrounds, which also means that the four stages are the way towards the fully fledged establishment of private property, since before the commercial society, governmental protection did not cover all private property.

People respect private property when they justify self-interested behavior. Protecting one's property requires the justification of pursuing one's own things. When an injury occurs, people recognize the proper object that they should protect from others. Smith wrote, "to disturb his happiness merely because it stands in the way of our own... is what no impartial spectator can go along with." Rather, he stated, "every man is, no doubt, by nature, first, and principally recommended to his own care; as he is fitter to take care of himself than of any other person, it is fit and right that it should be so." Because of the affirmation of taking care of oneself, "in the race for wealth, and honours, and preferments, he may run as hard as he can, and strain every nerve and every muscle, in order to outstrip all his competitors." However, "if he should juttle, or throw down any of them, the indulgence of the spectators is entirely at an end... they do not enter into that self-love by which he prefers himself so much to this other, and cannot go along with the motive from which he hurt him" (*TMS*, II. ii. 2. 1). As this suggests, self-interest as taking care of oneself indicates both the foundation and limit of the proper object of one's own. Because governments and people affirm self-interested behavior, they see things derived from pursuing wealth as one's proper property. At the same time, if one hurts another in the course of attaining wealth, the objects attained by it cannot be seen as one's proper things.

Certainly, this does not mean that Smith saw individuals solely as self-interested. He recognized that through imperfect rights, people can donate their own things. Nonetheless, these benevolent deeds are not compulsory, and do not constitute any part of jurisprudence. Thus, perfect and imperfect rights are connected to two kinds of justice: "a perfect right... relates to commutative justice. Imperfect rights... refer to distributive justice. The former are the rights which we are to consider, the latter not belong properly to jurisprudence, but rather to a system of moralls [sic] as they do not fall under the jurisdiction of the laws" (*LJ(A)*, i. 15). Smith regarded non-self-interested motives as relating to morals, whereas he saw self-interested motives as relating to jurisprudence. Again, Smith did not deny the existence of non-self-interested motives in human nature. Nonetheless, Smith regarded jurisprudence as requiring compulsory rights, and saw self-interested behavior as the object of jurisprudential consideration.

As a result, Smith regarded self-interested economic behavior as part of his explanation of the role of private property in the system of natural jurisprudence. He saw the protection of one's proper things or property as based on self-interested behavior. As I shall argue in the next section, Smith considered the commercial society, with its mechanism of wealth creation, as the first society to reconcile the full protection of property with the enjoyment of subsistence, the latter of which necessitated the violation of property rights in previous social stages. The commercial society makes possible the actual protection of one's proper things, or self-interested behavior.

### ***3.1 Self-interest in Economic Behavior***

This part examines how Smith described self-interest in economic behavior, and how self-interest was based on the jurisprudential framework of human behavior. Whereas people respect the use of justice to protect property, justice also illustrates the appropriate way of getting property. Indeed, Smith divided property into personal and real rights in *LJ*. He wrote that a real right is “a right to a particular thing;” whereas a personal right is “a right against a particular person” (*LJ(A)*, i. 16). About personal rights, he wrote, “a personal right may sometimes be constituted against a certain thing, but then it [is] only extends to that thing when in the possession of a certain person. Thus e. g. if a man sells another man a horse, but the horse is not delivered, the buyer has a personal right against the seller” (*LJ(A)*, i. 16). Real rights are categorized by dominium, or “the full right of property,” servitudes, pledge, and inheritance (*LJ(A)*, i. 16–19).

Indeed, when people engage in economic behavior, they can obtain property by protecting justice. They obtain property when they exchange goods. Basically, “man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favour, and shew [sic] them that it is for their own advantage to do for him that what he requires of them” (*WN*, I. ii. 2). As argued above, Smith saw benevolence, by which one gives another one's own things, as not being part of under the area of jurisprudence because jurisprudential justice depends on protecting one's property. The above quote presupposes this jurisprudential justice.

This protection rests on self-interested human nature because it does not presuppose benevolent human nature. Although Smith respected benevolence in human nature, he regarded people in exchange as self-interested because they want to obtain property. Protecting justice requires self-interest because people try to protect their own property. Indeed, Smith argued,

We obtain from one another the far greater part of those good offices which we stand in need of. It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love but to their self-love, and never talk to them of our own necessities but of their advantages (*WN*, I. ii. 2).

In trade situations, people try to gain the goods that others produce and sell. However, they cannot gain those goods for free; they have to pay for the goods they buy. This exchange presupposes obtaining one's own property while obeying justice. Jurisprudential justice depends on a self-interested human nature. People can exchange goods if they protect each other's property; they are not voluntarily donating their property for free. In this sense, this space excludes benevolence.

Smith made clear what the nature of the economic motive is. He said,

The preservation and healthful state of the body seem to be the object which Nature first recommends to the care of every individual...As he grows up, he soon learns that some care and foresight are necessary for providing the means of gratifying those natural appetites, of procuring pleasure and avoiding pain, of procuring the agreeable and avoiding the disagreeable temperature of heat and cold. In the proper direction of this care and foresight consists the art of preserving and increasing what is called his external fortune...The desire of becoming the proper objects of this respect, of deserving and obtaining this credit and rank among our equals, is, perhaps, the strongest of all our desires, and our anxiety to obtain the advantages of fortune is accordingly much more excited and irritated by this desire, than by that of supplying all the necessities and conveniences of the body, which are always very easily supplied (*TMS*, VI. i. 1–3).

People can obtain subsistence and property justly. The obtention was originally a way of satisfying our natural necessities but, in reality, was less important than earning people's reputation and rank. Because human basic necessities are "always very easily supplied," property also functions as the means of building a reputation. Property is based on the imagined human relationships; that is, the aim of ambition is to gain the attention of others; he said, "the wages of the meanest labourer can supply them [the necessities of nature]...he spends a great part of them [wages] upon conveniences, which may be regarded as superfluities" (*TMS*, I. iii. 2. 1). Ambitious people want to acquire a reputation in order to acquire more property.

Acquiring property effectively is seen as prudence: "the care of the health, of the fortune, of the rank and reputation of the individual, the objects upon which his comfort and happiness in this life are supposed principally to depend, is considered as the proper business of that virtue which is commonly called Prudence" (*TMS*, VI. i. 5). "CONCERN for our own happiness recommends to us the virtue of prudence." Prudence is "originally recommended to us by our selfish...affections" (*TMS*, VI. concl. 1). Prudence is different from self-interest and refers to a reasonable way of seeking "our own happiness" (*TMS*, VI. concl. 1). Self-interest itself was originally a matter of human intuition but was arranged to achieve its aim through prudence. Prudence is "our concern for our own happiness," as noted earlier. Prudence is self-regard, which arranges effectively one's own deeds and so is different from one's original selfishness.

Furthermore, self-interest is evaluated by others. Thrifty merchants are respected only when they behaved in a thrifty manner, i. e., they are respected not because they pay attention to trivial interests, but because they obey the general rule of thrifty conduct. Self-interest is socially evaluated as well as promoted by society. Overall, exercising self-interest means taking care of oneself and one's proper things. However, this self-interest requires the involvement of other people in order to build

a reputation; thus, self-interest is socially arranged, and the regard for others' reputation is the basic concern of the self. The self is formed through social interaction, as Smith's concept of sympathy suggests. Through sympathy, one reflects others' feelings, while others' feeling reflects one's own feelings through sympathy. Sympathy is the key to resolving a conflict between the self and others.

Although human self-interest is based on human relationships, self-interested behavior differs in each social stage. Smith said,

In the age of agriculture, they are not perhaps so much exposed to theft and open robbery, but then there are many ways added in which property may be interrupted as the subjects of it are considerably extended. . . . In the age of commerce, as the subjects of property are greatly increased the laws must be proportionally multiplied. The more improved any society is and the greater length the severall [sic] means of supporting the inhabitants are carried, the greater will be the number of their laws and regulations necessary to maintain justice, and prevent infringements of the right of property. (*LJ(A)*, i. 34–35)

*WN* aimed to examine the source of wealth in the commercial society where the extension of property meant that of one's proper things. Before the commercial society, in the agricultural society, that is "in a country which has neither commerce, nor any of the finer manufactures, a great proprietor, having nothing for which he can exchange the greater part of the produce of his lands which is over and above the maintenance of the cultivators, consumes the whole in rustick hospitality at home" (*WN*, III. iv. 5). In this agricultural stage, the demand and supply of goods were not smoothly exchanged, so voluntary donation or hospitality was needed. This hospitality was the source of the power of nobles, who could make their dependents wage war. Hospitality is a kind of benevolence that Smith admired in *TMS*. However, in his natural jurisprudence, where he considers the society of his day, benevolence was connected to authority and violence. This stage was full of war between nobles and kings. Instead, the commercial society provides nobles with products they can purchase, so they lose their authority. In commercial society, people enjoy the protection of their property and thus engage in self-interested behavior.

Thus, protecting property means that one should obtain one's necessities from someone whose property is also protected and thus the exchange of goods—not donation, hospitality, or deprivation—is necessary for mutual support. Only in the commercial society can the full protection of property be achieved because the people in the commercial society can subsist without using violence, deprivation, or donation. Reconciliation of the protection of one's proper things with subsistence takes the form of an exchange in the market, which can be fully achieved only in the commercial society. Before the division of labor, people could live off of their own products. However, in the commercial society, "when the division of labour has been once thoroughly established, it is but a very small part of a man's wants which the produce of his own labour can supply." Rather, "he supplies the far greater part of them by exchanging that surplus part of the produce of his own labour. . . . for such parts of the produce of other men's labour." As a result, "every man thus lives by exchanging, or becomes in some measure a merchant, and the society itself grows to be what is properly a commercial society" (*WN*, I. iv. 1). As this implies, only in the commercial society can people fully exchange their goods,

and people do not donate their “surplus part of the produce” for free. Instead, they try to sell the surplus; without selling it, they cannot survive. The mutual exchange in the commercial society presupposes mutual selling and buying, both of which are intended to acquire property while protecting one’s own property. Protecting property assumes self-interested individuals who do not donate their goods for free.

Because the commercial society is the model of the most developed society in *WN*, self-interest is the main motivator of conduct. Before the commercial society, subsistence was gained by some kind of benevolence connected with authority. The hospitality of the great landlords and churches in the Middle Ages were authoritative. Smith did not applaud them, because what he believed that benevolence should be voluntary and free (*TMS*, II. ii. 1. 3). In the commercial society only, one’s proper things are protected. For Smith, the basis of social maintenance was not to infringe upon others and to keep one’s proper things.

Overall, protecting one’s proper things depends on self-interest, which means acquiring material property as well as property as the means of earning a social reputation. Self-interest is the basic unit of human motive in Smith’s natural jurisprudence where justice presupposes not infringing on others’ proper things. Protecting property supposes that people are self-interested and justice as protecting one’s proper things demands self-interest in human relationships. Society is built on the protection of justice. The protection of property can be precarious because it can be easily violated. Only in the commercial society can the full protection of property and enjoyment of subsistence be achieved.

## 4 Conclusion

The apparent difference in the depictions of human nature in *TMS* and *WN* is derived from the natural jurisprudence of the latter. *WN* aims to explain the social mechanism of the full exercise of justice. Self-interest in *WN* is the means of acquiring wealth, which is a prerequisite for the maintenance of one’s proper things. By contrast, *TMS* includes not only self-interest and prudence but also benevolence and other superior virtues because it describes those virtues as recommendable for individuals and society. The two works assert the same human nature. However, in *WN*, protecting one’s proper things is the basic unit of the maintenance of society and justice. Protecting one’s proper things presupposes self-interest.

The difference in the descriptions of human nature in the two works is not a conflict between morality and economics. Economic behavior is based on a jurisprudential framework that aims to maintain society. *TMS* includes virtues both for the maintenance of society and for individual excellence, the former being the basis of human motivation in Smith’s natural jurisprudence including *WN*.

## References

- Brown, Vivienne. 1994. *Adam Smith's discourse; Canonicity, commerce, and conscience*. London and New York: Routledge.
- Force, Pierre. 2003. *Self-interest before Adam Smith: A genealogy of economic science*. Cambridge: Cambridge University Press.
- Göğmen, D. 2007. *The Adam Smith problem: Reconciling human nature and society in The Theory of Moral Sentiments and The Wealth of Nations*. London and New York: Tauris Academic Studies.
- Griswold Jr., C.L. 1999. *Adam Smith and the virtues of Enlightenment*. Cambridge: Cambridge University Press.
- Haakonssen, Knud. 1996. *Natural law and moral philosophy: From Grotius to the Scottish Enlightenment*. Cambridge: Cambridge University Press.
- Grotius, Hugo. [1625] 1901. *The rights of war and peace*, A.C. Campbell (trans. and ed.), David J. Hill (intro.), M. Walter Dunne, New York.
- Hobbes, Thomas. 1991. *Leviathan*, Richard Tuck (ed.). Cambridge: Cambridge University Press.
- Hume, David. 1985. Of the original contract. In *Essays, moral, political, and literary*, revised ed, ed. E.F. Miller. Indianapolis: Liberty Fund.
- Hutcheson, Francis. 1747. *A short introduction to moral philosophy*. Glasgow.
- Hutcheson, Francis. [1725] 2004. *An inquiry into the original of our ideas of beauty and virtue in two treatises*, ed. Wolfgang Leidhold. Indianapolis: Liberty Fund.
- Leonidas, Montes. 2004. *Adam Smith in context: A critical reassessment of some central components of his thought*. Basingstoke: Palgrave Macmillan.
- Nieli, Russell. 1986. Spheres of intimacy and the Adam Smith problem. *Journal of the History of Ideas* 47 (4): 611–624.
- Otteson, J.R. 2002. *Adam Smith's marketplace of life*. Cambridge: Cambridge University Press.
- Raphael, D.D., and A.L. Macfie. 1982. Introduction. In *Adam Smith, The theory of moral sentiments*, ed. D.D. Raphael and A.L. Macfie. Indianapolis: Liberty Fund.
- Pack, S.J. 1997. Adam Smith on the virtues: A partial resolution of the Adam Smith problem. *Journal of the History of Economic Thought* 19 (Spring): 127–140.
- Raphael, D.D. 2007. *The impartial spectator: Adam Smith's moral philosophy*. Oxford: Clarendon Press.
- Smith, Adam. [1759] 1979. *The theory of moral sentiments*, ed. D.D. Raphael and A.L. Macfie. Indianapolis: Liberty Fund.
- Smith, Adam. 1981. *An inquiry into the nature and causes of the wealth of nations*, ed. R.H. Campbell, A.S. Skinner, and W.B. Todd. Indianapolis: Liberty Fund.
- Smith, Adam. 1982. *Lectures on jurisprudence*, ed. R.L. Meek, D.D. Raphael, and P.G. Stein. Indianapolis: Liberty Fund.
- Sonenscher, Michael. 2009. Ideology, social science and general facts in late eighteenth-century French political thought. *History of European Ideas* 35 (1): 24–37.
- Tierney, Brian. 2001. *The idea of natural rights: Studies on natural rights, natural law, and church law, 1150–1625*. Michigan: Wm. B Eerdmans Publishing Co.
- Tuck, Richard. 1979. *Natural rights theories: Their origin and development*. Cambridge: Cambridge University Press.
- Viner, Jacob. 1960. The intellectual history of laissez faire. *Journal of Law and Economics* 3: 45–69.
- Young, Jeffrey T. 1997. *Economics as a moral science: The political economy of Adam Smith*. Cheltenham and Lyme: Edward Elgar.



# Bentham on Self-interest: Institutional Control of Self-interest



Hiroaki Itai

**Abstract** I examine Jeremy Bentham on self-interest. The originality of Bentham's utilitarianism seems clear, and he was not particularly concerned with the problems of selfishness. He presupposes the existing social order, and assumes that it does not matter whether human beings are selfish or altruistic, since, as he himself acknowledged, a utilitarian system of governance transforms people into agents of appropriate utilitarian accounting and maximizes social well-being. While people are free to the extent that they do not harm others, delinquency and misrule, which impede the growth of social well-being, are controlled both through oversight, and through physical mechanisms such as the Panopticon, by legislators, and the general public. In this sense, Bentham may have changed the liberal tradition's way of considering selfishness.

**Keywords** Utilitarianism · Hedonism · Publicity · Enlightenment · Surveillance

## 1 Introduction

In this chapter, I examine Jeremy Bentham on self-interest. Bentham views 'self-interest' as a theoretical word and 'selfish' as a contempt word; the two terms will be considered as different concepts in this chapter. The goal is to clarify the originality of Bentham utilitarianism. Bentham's discussion of how the selfishness/self-interest discourse unfolded between the 18th and nineteenth centuries will shed light on this.

First, I shall take a quick look at how Bentham has been positioned in the history of social philosophy. What has traditionally been emphasized in treatments of Bentham's utilitarianism is the secular vision of his philosophy: he rejected natural law, natural rights, and the system of the Church of England; he considered utilitarianism to be grounded in empirical findings about pleasure and pain.

Unlike the philosophy of Hume or Smith, Bentham's utilitarianism was a philosophy that provided a framework for social engineering and was indifferent to the

---

H. Itai (✉)

Institute for Gender Studies, Ochanomizu University, Tokyo, Japan

e-mail: [brahms@joy.hi-ho.ne.jp](mailto:brahms@joy.hi-ho.ne.jp)

© Springer Nature Singapore Pte Ltd. 2021

S. Egashira et al. (eds.), *A Genealogy of Self-Interest in Economics*,  
[https://doi.org/10.1007/978-981-15-9395-6\\_5](https://doi.org/10.1007/978-981-15-9395-6_5)

nature and origin of moral sentiments.<sup>1</sup> Bentham considered himself to have inherited Hume's criticism of social contract theory and, in distinguishing 'to be' from 'ought to be', to show the influence of Helvétius in legislative theory, Beccaria in penal theory, Locke and d'Alembert in linguistic theory, and Smith in economics. In this sense, Bentham is considered to have a syncretistic rather than a unique philosophy.<sup>2</sup>

Since the turn of the century, Kelly's (1999) interpretation of Bentham as a member of the Republican lineage in the development of the constitutional theory of government, and Rosen's (2003) study positioning Bentham as a member of the tradition of 'modern Epicureanism' in the development of social well-being have been influential. Kodama (2010) attempted to draw up a history of utilitarianism centring on Bentham from the viewpoint of the conflict between 'intuitionism' and 'utilitarianism' by referring to Schneewind. While these arguments can certainly be evaluated as attempts to position Bentham's utilitarianism within the history of philosophy, they do not seem to emphasize the uniqueness of Bentham's utilitarianism. The following discussion of his concept of self-interest, which does bear some unique features such as controlling self-interest by some external means helps us better understand some of the reasons why Bentham is difficult to position in the history of thought.

## 2 Hedonistic Human Conduct in Bentham's Utilitarianism

Bentham's view of human beings is hedonistic. On the first page of *An Introduction to the Principles of Morals and Legislation* (1996, hereafter *IPML*), he says, 'Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne'.<sup>3</sup> Pleasure controls all of human behaviour, speech, and thought, and thus right and wrong are also closely connected with pleasure and pain:<sup>4</sup> 'The only object of man is to seek pleasure and to avoid pain, even at the very moment when he rejects the greatest pleasure or accepts the greatest

---

<sup>1</sup>See Bentham (1996), Chaps. 1–3; see also Crimmins (1990). However, some point out that it is necessary to reserve judgement on whether or not Bentham's criticism of religion is a criticism of Christianity as a whole (Obata 2015).

<sup>2</sup>Plamenatz (1949). Bentham's 'On Utilitarianism' tells us which thinkers in the past overlapped with Bentham's thoughts. In this late article, Bentham came to admit that the term utilitarianism was inappropriate as an expression of his philosophy, but for the image of Benthamism in the nineteenth century as 'demon' (T. Arnold) and 'Pig philosophy' (T. Carlyle), see Izumiya (1976).

<sup>3</sup>Bentham (1996), Chap. I, 1.

<sup>4</sup>Bentham makes a strict distinction between 'good and bad' and 'right and wrong'.

pain'.<sup>5</sup> Even if someone is hungry, distributing food to impoverished neighbours is an act of enjoyment, not a denial of pleasure.<sup>6</sup>

Also, because pleasure can be felt by anyone, all sentient beings, including animals, should be treated equally.<sup>7</sup> 'Pain and pleasure at least, are words which a man has no need, we may hope, to go to a Lawyer to know the meaning of'.<sup>8</sup> The equality in comprehensibility of pleasure and pain leads to equality between human beings regardless of rank or property. An atomistic<sup>9</sup> individual, stripped of social relations, is thus Bentham's theoretical starting point. It is only pleasure and pain that drives individuals into action; let us now see how the individual is affected by them.

## 2.1 Behaviour and Sanction

I examine Bentham's understanding of human behaviour in accordance with the IPML written as the introduction to the principles of the Criminal Code. First, human behaviour begins from some 'exciting cause',<sup>10</sup> and pleasure and pain, and with them any sanctions, become 'motives' ('prospect of pleasure and pain'); broader 'dispositions' are then formed from a human being's motives.<sup>11</sup> Behaviour is the result of going through various circumstances, such as time, place, and human relations, and may emerge due to 'consciousness, unconsciousness [or] mistake', or some combination of these.<sup>12</sup> The *result* of a behaviour is the production and the subsequent reception of pleasure and/or pain, which constitutes the 'material.'<sup>13</sup> It is only after

---

<sup>5</sup>Bentham (1931), p. 2.

<sup>6</sup>Bentham's use of the physical in both sense and soul (Bentham 1931, p. 3.) is suggestive of his interpretation of physical sanction.

<sup>7</sup>Bentham (1931), p. 3.

<sup>8</sup>Bentham (1977), p. 418.

<sup>9</sup>Bentham (1996), Chap. I, 4.

<sup>10</sup>Bentham (1996), Chap. VI, 4.

<sup>11</sup>Bentham (1996), Chap. XI, 28.

<sup>12</sup>Bentham (1996), Chap. XII, 7, Ch. III. Bentham stated that 'the nature of a man's disposition must depend upon the nature of the motives he is apt to be influenced by,' and 'the disposition he is of during a certain period, the sum or result of his intentions during that period.' (Bentham 1996, Chap. IX, 27). About intention: 'The exciting cause, the pleasure or pain produced by it, and the intention produced by such pleasure or pain in the character of a motive, are objects so intimately connected, that, in what follows, I fear I have not, on every occasion, been able to keep them sufficiently distinct.' (Bentham 1996, Chap. VI, 4n(a)). About motive: 'a motive must be the prospect of some pleasure, or other advantage, to be enjoyed in future.' (Bentham 1996, Chap. XII, 7). Bentham categorizes motives into 'a motive in prospect' and 'a motive in essence' (Bentham 1996, Chap. X, 6).

<sup>13</sup>According to Bentham, 'materiality is a relative term: applied to the consequences of an act, it bore relation to pain and pleasure: applied to the circumstances, it bears relation to the consequences.' (Bentham 1996, Chap. VII, 23). He analysed this 'materiality' in Chap. VII et seq. of *IPML*. Before Chap. 7, the only example of 'substantivity' is the material in Chap. III (Bentham 1996, Chap. III,

the consequences of a behaviour are seen that the tendency of a motive or action is shown to be either right or wrong.<sup>14</sup>

Thus, the cause of human conduct is a sanction based on pain or pleasure, which is derived from an origin that is physical, political, moral, or religious.<sup>15</sup> The other three types all emerge from the physical sanction.<sup>16</sup> The motives, pleasures, and principles for responding to the four sanction types are as follows which Bentham refers to as:

1. *physical sanction*, entailing ‘the motive of benevolence’,<sup>17</sup> ‘the pleasures of goodwill, the pleasures of sympathy, or the pleasures of the benevolent or social affections’,<sup>18</sup> or ‘the principle of benevolence’;<sup>19</sup>
2. *political sanction*, entailing ‘the motive of self-preservation’,<sup>20</sup> ‘the motives, whereof the influence is at once most powerful, most constant, and most extensive, are the motives of physical desire, the love of wealth, the love of ease, the love of life, and the fear of pain: all of them self-regarding motives’,<sup>21</sup> or ‘the principle of self-interest’;<sup>22</sup>
3. *moral sanction*, entailing ‘the fear of shame’,<sup>23</sup> ‘the pleasures of a good name’, ‘the love of reputation’,<sup>24</sup> ‘the pleasures of good repute, the pleasures of honour’,<sup>25</sup> ‘the pains of ill-repute, the pains of dishonour’,<sup>26</sup> or ‘the principle of love of reputation’;<sup>27</sup>

---

5n (b)), which was added in January 1789, according to the Editor’s Note to IPML (Bentham 1996, Chap. III, 5n (2)).

<sup>14</sup>Bentham (1996), Chap. X, 12; Chap. XII, 1; Chap. VI.

<sup>15</sup>‘inasmuch as the pleasures and pains belonging to each of them [distinguishable sources] are capable of giving a binding force to any law or rule of conduct, they may all of them be termed sanctions’ (Bentham 1996, Chap. III, 2). It should be noted that sanctions here can cause not only pain but also pleasure. The sanction is a concept within the ‘resource’ that works on both pleasure and pain and influences behaviour.

<sup>16</sup>Physical sanctions have two meanings: physical action and action based on human nature (Dinwiddy 1989, p. 23). In his later *Fragment on Ontology* (1813, 1814, 1821), in the Appendix of his nomography paper on the drafting of the law (1811–13) (Bentham 1838–43, iii, pp. 285–295), and in *A Table of the Springs of Action* (Bentham 1983, p. 8), sanctions were classified into five categories: 1. physical sanctions, 2. sympathetic sanctions or sympathy sanctions, 3. popular or moral sanctions, 4. political sanctions, including legal sanctions, and 5. religious sanctions. The physical element and the element inherent in human nature, which were included in physical sanctions in *IPML*, were thus separated into two.

<sup>17</sup>Bentham (1996), Chap. XII, 8.

<sup>18</sup>Bentham (1996), Chap. V, 10. See also Chap. V, 26.

<sup>19</sup>Bentham (1996), Chap. XI, 20.

<sup>20</sup>Bentham (1996), Chap. XII, 8.

<sup>21</sup>Bentham (1996), Chap. XII, 33.

<sup>22</sup>Bentham (1996), Chap. XI, 20.

<sup>23</sup>Bentham (1996), Chap. XII, 8.

<sup>24</sup>Bentham (1996), Chap. X, 22.

<sup>25</sup>Bentham (1996), Chap. V, 7.

<sup>26</sup>Bentham (1996), Chap. V, 24.

<sup>27</sup>Bentham (1996), Chap. XI, 20.

4. *religious sanction*, entailing ‘the fear of the divine displeasure’,<sup>28</sup> and ‘the pains of piety’.<sup>29</sup>

Bentham’s use of ‘self-interest’ as a principle of political sanction is historically, an extremely rare use of this term, and in *IPML* he generally refers to ‘self-regarding principles’. These sanctions corroborate and influence human behaviour,<sup>30</sup> and are classified and understood as physical, political, moral, or religious sanctions from the external perspective of social binding as well as from the internal perspective of personal binding.

## 2.2 *Self-interest and Other-Interest*

Bentham sees human beings as pursuing pleasure and avoiding pain, but he does not see them as selfish as selfish in the sense of negatively self-oriented. According to a note added by Bentham in July 1822 to the second edition of the *IPML*, he changed what was formerly ‘the principle of utility’ to ‘the principle of greatest happiness,’ because ‘[t]he word utility does not so clearly point to the idea of pleasure and pain as the words happiness and felicity do: nor does it lead us to the consideration of the number, of the interests affected; to the number, as being the circumstance, which contributions, in the greatest proposition, to the formation of the standard here in question’.<sup>31</sup>

In the *Table of Action* (1815), it is Bentham who acknowledges that people generally perceive the principles of utility and utilitarianism as selfish, whereas Bentham advocates the principle of greatest happiness for the greatest number, arguing that utilitarianism is not ‘selfish’.<sup>32</sup> In *Emancipate Your Colonies* (1820), Bentham again

<sup>28</sup>Bentham (1996), Chap. XII, 8.

<sup>29</sup>Bentham (1996), Chap. V, 25, Chap. V, 9.

<sup>30</sup>Bentham argued that ‘the political sanction, ..., is but one of four sanctions, which may all of them contribute their share towards producing the same effects’ (Bentham 1996, Chap. XIV, 26) and that the power of sanctions other than political sanctions should be considered. On competition from sanctions: ‘1. The motive of benevolence, which acts as a branch of the physical sanction. 2. The motive of self-preservation, as against the punishment that may stand provided by the political sanction. 3. The fear of shame: a motive belonging to the moral sanction. 4. The fear of the divine displeasure: a motive belonging to the religious sanction. On the first and last of these forces it has, perhaps, no influence worth insisting on: but it has on the other two’ (Bentham 1996, Chap. XI, 20., Chap. XII, 8).

<sup>31</sup>Bentham (1996), Chap. I, 1n(a).

<sup>32</sup>Bentham (1983), pp. 56–57. The ‘principle of general utility’ is also used in *A Table of the Springs of Action*, where it is stated that there are two types of principle of utility: ‘the censorial and the expositive or exhibitiv’ (Bentham 1983, p. 9). As for ‘censorial’ and ‘expositive,’ Bentham pointed out in his *Fragment on Government*: ‘To the province of the Expositor it belongs to explain to us what, as he supposes, the Law is: to that of the Censor, to observe to us what he thinks it ought to be.’ Bentham (1977), p. 397.

used this word with negative connotations,<sup>33</sup> saying that it was ‘selfish’ not to give up the Spanish colonies. Thus we can see that Bentham pays keen attention to the difference between selfishness and utilitarianism.

In fact, the self-interested practices that Bentham refers to do not distinguish self-interest from other-interest. The 14 pleasures Bentham classifies in *IPML* include those of fellowship, piety, and charity alongside those of feeling and wealth<sup>34</sup>. The pleasure of mercy,<sup>35</sup> which stimulates the act of mercy described as altruistic, is the pleasure obtained by the act of mercy towards others. The act which brings pleasure to the self is not limited to merely benefiting the self, and the person who pursues pleasure gains it not only for reasons confined to himself but also as pleasure rooted in relationships with others.<sup>36</sup> Here, the selfish/altruistic dichotomy has receded, and in general, a self-seeking behaviour in Bentham’s utilitarianism is not selfish or altruistic, but self-regarding or extra-regarding.

### 2.3 *Extra-Regarding Motives and Universal Interest*

To examine the utilitarian concept of caring for others in utilitarianism, I consider the three rules of moral obligation that Bentham set forth to distinguish between the legislation responsible for public policy and the private ethics responsible for individual ethical conduct. The first of the rules of moral obligation is ‘prudence’,<sup>37</sup> and the legislation’s interference with human desires is the most unnecessary.<sup>38</sup>

Second, there is a rule of moral obligation, probity, ‘to discharge the negative branch of it’, that is, ‘forbearing to diminish’ the happiness of one’s neighbourhood

---

<sup>33</sup>Bentham (1995), p. 344. Bentham states that there is no system of commercial policy that ultimately benefits solely on the basis of selfish principles in *Restricted and Prohibited Commercial System Considerations* (1820, Bentham 1995, p. 347). He also used ‘selfish tyranny’ to describe the sinister interests of the ruling class in a monarchy in *Codification and Public Education* (Bentham 1998, p. 56).

<sup>34</sup>Bentham (1996), Chap. V. According to Bentham, the pleasures of seeing the countryside consist of: I. Pleasures of the senses: 1. The simple pleasures of sight, 2. The simple pleasures of the ear, 3. The pleasures of the smell, 4. The agreeable inward sensation; and II. Pleasures of the imagination produced by association: 1. The idea of the plenty, 2. The idea of the innocence and happiness of the animals, 3. The idea of the constant flow of health, 4. The idea of gratitude (Bentham 1996, Chap. V. 33n(p)). In *A Table for Springs of Action*, Bentham distinguishes ‘derivative’ pain and pleasure caused by memory and imagination from ‘original’ pain and pleasure caused by physical and psychological sensations (Bentham 1983, p. 10).

<sup>35</sup>Bentham (1996), Chap. V, 10.

<sup>36</sup>Of course, even when a person acts based on the extra-regarding motive of consideration for others, the pursuit of increasing his pleasure is fundamental. Such pleasures include ‘malice pleasures’ (Bentham 1996, Chap. V, 2). Therefore, the principle of utility is not merely an increase in pleasure, but an increase in happiness as well (a state in which pleasure exceeds pain).

<sup>37</sup>Bentham (1996), Chap. XVII, 6.

<sup>38</sup>‘It is plain, that of individuals the legislator can know nothing’ (Bentham 1996, Chap. XVII, 15). Bentham’s argument that each person is the best judge of their interests and that legislation should not interfere can be found in *Defense of Usury*. See Senga (1988).

is part of the ‘the art of discharging one’s duty to one’s neighbour’.<sup>39</sup> This is with legislative assistance indispensable<sup>40</sup> in terms of the ‘benefits for punishing those who harm their neighbours’ with respect to remedies for property infringement and approval of property.<sup>41</sup> Third, the other ‘the art of discharging one’s duty to one’s neighbour’ is the rule of ‘discharg[ing] the positive branch, beneficence’,<sup>42</sup> which is ‘intended to promote the happiness of neighbours’. Bentham points out that, for example, when a drunken man is suffocating in a puddle, someone does not rescue him<sup>43</sup>; there is thus room for legislation to intervene with respect to the rules of beneficence, which means not to increase one’s own happiness but rather to increase the happiness of others. Bentham stated that ‘the only interests which a man at all times and upon all occasions is sure to find adequate motives for consulting, are his own’, but that ‘notwithstanding this, there are no occasions in which a man has not some motives for consulting the happiness of other men. In the first place, he has, on all occasions, the purely social motive of sympathy or benevolence: in the next place, he has, on most occasions, the semi-social motives of love of amity and love of reputation’.<sup>44</sup>

Bentham postulates in *IPML*<sup>45</sup> that ‘happily there is no primeval and constant source of antipathy in human nature, as there is of sympathy’ and in *Deontology* (published 1834)<sup>46</sup> that beneficence is driven by (1) expectation of rewards (retributive sanction), (2) expectations for publicity and praise (moral sanction), and (3) motivations for empathy (sympathetic sanction). Just as Bentham organized the motivations of human behaviour into self-regarding motives, antisocial motives, and social motives, people have the motivation to engage in social behaviour.<sup>47</sup> However, this is a different matter from the inherent propensity towards beneficence and sympathy. In fact, human beings are social beings receiving the cultural influence from their own society; according to the Bentham, it is human beings, for instance, whose prejudice

<sup>39</sup>Bentham (1996), Chap. XVII, 6.

<sup>40</sup>‘We must first know what are the dictates of legislation, before we can know what are the dictates of private ethics.’ (Bentham 1996, Chap. XVII, 18) see also Bentham (1838–43), I, pp. 155–168.

<sup>41</sup>Bentham (1983), p. 154.

<sup>42</sup>Bentham (1996), Chap. XVII, 6.

<sup>43</sup>Bentham (1996), Chap. XVII, 19n. In this case there is a moral obligation to prevent the loss of happiness of others. ‘Idleness’ is ‘the mother of vice and misery’ (Bentham 1996, Chap. II, 8), and ‘industry’ is encouraged in *Panopticon* (Bentham 1838–43, iv, p. 44).

<sup>44</sup>Bentham (1996), Chap. XVII, 7.

<sup>45</sup>Bentham (1996), Chap. VI, 27.

<sup>46</sup>Bentham (1983), pp. 182–183. Since Bentham considers private ethics as an individual’s ethical behaviour in *Deontology*, he considers the virtue of actors who are hardly questioned in the field of legislation. The virtue of beneficence is positioned slightly different from *IPML*. First, Bentham tailed beneficence as a set of benevolence. He stated that ‘benevolence and beneficence are either positive or negative. Positive beneficence is doing good; negative beneficence is abstaining from doing evil’ (Bentham 1983, p. 183). Regarding the relationship between virtue and happiness in *Deontology*, see Kodama (1999).

<sup>47</sup>Bentham (1996), Chap. X, 34. Bentham classifies good-will as a purely-social motives, while lust for ‘the love of reputation, the desire of amity, and the motive of religion’ are classified as semi-social motives (Bentham 1996, Chap. X, 35).

justifies the punishment of homosexual acts, and the repression of women, which are harmless in society. For Bentham, there was no justification for these repressions.

Let us look at the uniqueness of Bentham's concept of interest in its relation to self-regarding and extra-regarding acts. Bentham refers to *interest* as 'one of those words, which not having any superior genus, cannot in the ordinary way be defined'<sup>48</sup> and says that 'a thing is said to promote the interest, or to be for the interest, of an individual, when it tends to add to the sum total of his pleasures: or, what comes to the same thing, to diminish the sum total of his pains'.<sup>49</sup> For Bentham, interest would be a term that means almost the same thing as utility, as something that promotes increased pleasure and tends to decrease suffering.

It should be noted that Bentham used the term interest in a very broad sense, covering not only self-regarding interest, but also the interests he called extra-regarding or social.<sup>50</sup> And yet, Bentham used a rather singular concept of interest: 'If, however, instead of the word *happiness*, the word *interest* is employed, the phrase *universal interest* may be employed as corresponding indifferently to the interest of the greatest number, or to the interest of all'.<sup>51</sup> Engelmann points out that Bentham sublimated the concepts of national reason and public interest into one of 'monistic interest'<sup>52</sup> and regarded the aim as 'maximiz[ing] the felicity of *every one* of the individuals, of whose interests the universal interest is composed; on which supposition, the greatest happiness of *all*, not of the greatest number only, would be the end aimed at' as the purpose of governance.<sup>53</sup>

Bentham's concept of interest must be understood in terms of 'a clear view of [one's] own interest'.<sup>54</sup> Since people are easily perplexed by various prejudices, it is the intention of Bentham's utilitarianism appropriately to grasp the state of society and apply the principle of utility (the principle of greatest happiness) in a principled and coherent manner, based on 'reason and public utility'.<sup>55</sup> Bentham does not believe that legislators such as the 'Enlightened Monarch' achieve social welfare; more generally, Bentham sees people as self-regarded and does not believe that they always work to achieve social well-being.

---

<sup>48</sup>Bentham (1996), Chap. I, 5n(c).

<sup>49</sup>Bentham (1996), Chap., I, 5.

<sup>50</sup>Dinwiddy (1992), p. 433. See also Dinwiddy 1989, p. 23.

<sup>51</sup>Bentham (1838–43), IX, p. 6.

<sup>52</sup>Engelmann (2003), p. 3.

<sup>53</sup>Bentham (1838–43), II, p. 269n. 'The right and proper end of government in every political community, is the greatest happiness of all the individuals of which it is composed, say, in other words, the greatest happiness of the greatest number. ... The *actual* end of government is, in every political community, the greatest happiness of those, whether one or many, by whom the powers of government are exercised.' Bentham (1838–43), IX, p. 5.

<sup>54</sup>Bentham (1996), Chap. IV, 8.

<sup>55</sup>Bentham (1977), p. 402. Lyons interpreted the principle of utilitarianism as an individual, concrete principle (Lyons 1973), but Dinwiddy criticized this interpretation, arguing that it was a universal principle (Dinwiddy 1982, pp. 402–422).



### 3 Institutional Control of Selfishness

One way to control self-interest is to cultivate human nature, including conscience, virtue, and public spirit. However, Bentham clearly had the opinion that systems outside the human body, such as governance structures, may direct the expression of self-interest.

In this regard, there is a line of argument that Bentham respected individual decision-making as much as possible and defended freedom because ‘no man can be so good a judge as the man himself, what it is gives him pleasure or displeasure’.<sup>56</sup> But this interpretation requires caution; in fact, Bentham’s words merely point out that no one but oneself can feel one’s own pleasure and that no one else can judge from what it is desirable for that individual to obtain pleasure. Bentham also says, ‘A man may suffer grievously, for instance, by a new tax, without being able to trace up the cause of his sufferings to the injustice of some neighbour, who has eluded the payment of an old one’.<sup>57</sup> In this way, self-interest requires not only intuition on the part of the person concerned but also reasoning in light of a variety of circumstances. It is the role of the ruler to make such decisions and to establish a legal framework to enable people to make appropriate judgements on their own interests. This section examines how the ruler leads the people and also examines how people are not merely objects of manipulation by the ruler but have an active role in governing, as they understand themselves best.

#### 3.1 *Free State and Rational Discussion*<sup>58</sup>

I will examine what Bentham said about the ways in which the governed people were actively involved in governing and discuss three of his six conditions of ‘free state’<sup>59</sup> from his *Fragment on Government* published in 1776. The first condition is ‘the responsibility of the governors.’ By this he meant that the governor or legislator should make public the reasons behind his policies so that the governed can assess them. In such a way, a reasoning or accounting of opinions is made more efficient by way of critique and discussion about governmental actions. Second is ‘the liberty of the press’, defined as ‘the security with which every man, be he of the one class or the other, may make known his complaints and remonstrances to the whole

<sup>56</sup>Bentham (1996), Chap. XII, 4(1).

<sup>57</sup>Bentham (1996), Chap. II, para 15. ‘Every nation is liable to have its prejudices and its caprices, which it is the business of the legislator to look out for, and to cure’; in Bentham (1996), Chap. XV, 24. We emphasize that Bentham does not merely aggregate happiness across society. He recognizes so-called ‘adaptive preference formation’ (Elster 1982, pp. 219–238). As Bentham says, ‘The people have been quiet and submissive enough, it is true, under these several theocracies: but have they been happy? I should suppose not: if abject slavery, vain terrors, useless obligations, and tormenting self-denials are any obstacles to happiness.’ (Bentham MS, lxxxvii, 98).

<sup>58</sup>the contents of this and subsequent sections have been based and revised to Itai (1998).

<sup>59</sup>Bentham (1977), p. 485.

community'.<sup>60</sup> Public discussion can be meaningful only when the governed assert their opinion or bring to light the social problems of society in question. Third is '*the liberty of public association*', defined as 'the security with which malecontents may communicate their sentiments, concert their plans, and practise every mode of opposition short of actual revolt, before the executive power can be legally justified in disturbing them'.<sup>61</sup> The liberty of public association is important both as the objective condition for the formation of clubs or associations and as the condition for their members to present their opinions to legislators.

Based on these assumptions, Bentham's idea of society as 'a habit of reasoning and discussion [that] will penetrate all classes of society'<sup>62</sup> assumes concrete form. In short, government must always be open to the criticism of the governed. Also, because 'a system that is never to be censured, will never be improved',<sup>63</sup> 'every system of law', Bentham says, 'then may from time to time be requiring alterations'.<sup>64</sup> The active involvement of the people is therefore essential in governance.

Bentham's idea of government by rational and critical discussion is related to his principle of utility, for this is the basis of critical discussion that enables a grounding in 'fact,' and brings clarity to points of dispute.<sup>65</sup> The principle of utility is a 'principle of harmony' and makes possible such modes of rational discussion.<sup>66</sup> The principle of utility possesses an 'external standard' grounded in objective reality, and it is also a means for solving various problems of government and morality through reasonable discussion based on felicific calculus and fact.<sup>67</sup>

What is the subjective condition for society in which the behaviour of reasoning and discussion will penetrate all classes? For Bentham, it is for the good citizen to adopt the motto of '*To obey punctually; to censure freely*'. Strict obedience to the law and free criticism, by citizens are necessary for good governance. Moreover, Bentham supposes that every person can judge their own interest in applying the principle of utility.<sup>68</sup> He writes, 'for no man can be so good a judge as the man himself, what it is gives him pleasure or displeasure'.<sup>69</sup> Elsewhere, he writes that 'pain and pleasure at least, are words which a man has no need, we may hope, to go to a Lawyer to know the meaning of'.<sup>70</sup>

In Bentham's utilitarianism, critical behaviour is compatible with a habit of obedience according to the rules of a given political society. He also displays a belief in

---

<sup>60</sup>Bentham (1977), p. 485.

<sup>61</sup>Bentham (1977), p. 485.

<sup>62</sup>Bentham (1998), p. 31, Bentham (2002), pp. 408–9.

<sup>63</sup>Bentham (1977), p. 399.

<sup>64</sup>Bentham (1970), p. 236.

<sup>65</sup>Bentham (1977), p. 492.

<sup>66</sup>Nagai (1982), p. 9, Fukada (1984), p. 36, Burns (1974), p. 697.

<sup>67</sup>Bentham (1996), p. 25.

<sup>68</sup>Bentham (1776), p. 399.

<sup>69</sup>Bentham (1996), p. 159.

<sup>70</sup>Bentham (1776), p. 418.

the avoidance of ‘tyranny or anarchy’.<sup>71</sup> Related to this point, I will now discuss a fundamental framework of Bentham’s jurisprudence.

For most of his life, Bentham intended to reform the law system in England by his ‘codification’ because of his discontent with the English common law system. He felt that the system was as irrational as ‘the dark Chaos’,<sup>72</sup> that it oppressed a majority of the governed and that it privileged a small minority.<sup>73</sup> Bentham’s fundamental principle of ‘the greatest happiness of the greatest number’ should be understood in this context.

Therefore, the most important thing for Bentham was ‘*universal jurisprudence*’: methods on how to define legal terms such as rights and obligations that functioned in all legal branches and was ‘considered in respect of its form; in other words, in respect of its method and terminology’.<sup>74</sup> Every code based on universal jurisprudence must have a rational foundation of justification not only to compel the governed to obey but also so that the codes themselves can be reformed constantly. To this end, Bentham argued that there are ‘two parallel and connected systems, running on together, the one of legislative provisions, the other of political reasons, each affording to the other correction and support’.<sup>75</sup> The principle of utility as a tool of ‘*rational legislation*’<sup>76</sup> bears these systems and attempts to ‘rear the fabric of felicity by the hands of reason and of law’.<sup>77</sup>

### 3.2 *Reform and Mutual Enlightenment*

The project of reforming the common law system was thus to be achieved by a rational legal system made by the legislator and the critical behaviour of the governed. The first step in this process was the promulgation of law so that the governed could have a free discussion about political, legal, and social problems related to construction of social institutions. Bentham stated in his ‘*Essay on the Promulgation of the Law*’ that the governed would dutifully carry out lawful acts based on the promulgation of the law. He wrote, ‘To promulgate a law, is to present it, to the minds of those who are to be governed by it in such manner as that they may have it habitually in their memories, and may possess every facility for consulting it, if they have any doubts respecting what it prescribes’.<sup>78</sup> Here we see Bentham’s shift from an emphasis on obedience by the governed to an affirmation of their critical behaviour.

---

<sup>71</sup>Bentham (1999), p. 20.

<sup>72</sup>Bentham (1776), p. 198.

<sup>73</sup>Mill (1838), p. 145.

<sup>74</sup>Bentham (1996), p. 6. See also pp. 294–295.

<sup>75</sup>Bentham (1996), p. 9.

<sup>76</sup>Bentham (1998), p. 102.

<sup>77</sup>Bentham (1996), p. 11.

<sup>78</sup>Bentham (1838–43), I, p. 157

Bentham stated that ‘obedience to the laws would come to be hardly distinguishable from the feeling of liberty’.<sup>79</sup> According to Long, Bentham also stated that ‘the most fundamental characteristic of the idea of liberty, here as elsewhere in [his] writings, is its employment as a means of obtaining security, and its repudiation as an end in itself’.<sup>80</sup> At the same time, the promulgation of law would secure the expectations of the governed regarding the protection of property and life. As Bentham noted, ‘there are laws conformable to expectations already formed, laws which rest upon *natural expectations*’.<sup>81</sup> The relation in which the governor and the governed participate in critical discussion over public matters is what I want to call ‘mutual enlightenment’, or ‘network of mutual enlightenment’.<sup>82</sup> Under such conditions, information flows between the governed and the governing, providing political energy and sustenance to rational discussion.

However, the conditions of publicity that enable ‘mutual enlightenment’ can function as well as conditions of ‘mutual discipline’. In the next section, using the idea of ‘mutual discipline’, I will argue that an opportunity for subjection, and not only criticism, can be observed within ‘the motto of a good citizen’.

### 3.3 *Dependence, Control, and Mutual Discipline*

Bentham’s *Panopticon*,<sup>83</sup> written in 1787 in Russia, symbolically discussed the problem of the subjection of the governed. The panopticon aimed to improve prisoner treatment through efficient surveillance, both spatially and temporally, and by achieving the saving through a contract system.

Foucault’s splendid description of an essential characteristic of the panopticon is that ‘the panopticon is a machine for dissociating the see/being seen dyad; in the peripheric ring, one is totally seen, without ever seeing; in the central tower, one sees everything without ever being seen. It is an important mechanism, for it automatizes and disindividualizes power’.<sup>84</sup> The panopticon divides the surveyor from the surveyed through a transparent architectural layout in which the former is

<sup>79</sup>Bentham (1838–43), I, p. 161.

<sup>80</sup>Long (1977), p. 206. Bentham uses ‘liberty’ here in two different senses. One is the sense of civil and political liberty as produced by positive law. The other is the sense of liberty as a natural right which potentially results from anarchism (Rosen 1992, pp. 25–39).

<sup>81</sup>Bentham (1931), p. 150., Bentham (1931), pp. 96–8.

<sup>82</sup>Gaonkar (1994), p. 555. This is the same as Habermas’s understanding of publicity or public opinion in late eighteenth-century Britain. Habermas (1990), SS. 174–5. Habermas, citing Fox’s 1792 speech in the House of Commons, assumes the function of public criticism to operate by the official, but indirect, recognition of political affairs by a public who discusses political matters (Habermas (1990), S. 131). See also Parliamentary (1966), p. 974.

<sup>83</sup>The idea of the panopticon is hinted at in educational and military facilities established in the eighteenth century. Bentham also aligned himself with the intellectual world of the eighteenth century with respect to his notion of a transparently ruled space (Bender (1987), pp. 2–5).

<sup>84</sup>Foucault (1975), pp. 201–202.

made invisible and the latter is made completely visible.<sup>85</sup> A symmetrical relation of ‘seeing = being seen’ is converted to an asymmetric relation of ‘*seeing without being seen*’.<sup>86</sup> I see this asymmetric power as that which secures absolute submission in Bentham’s *Panopticon*.

The prisoner located in the outer ring of panopticon is made a suitable subject for civil society by the surveillance tower at the centre. The prisoner voluntarily learns some codes of industry suited to civil society and, as a result, forms a self voluntarily Foucauldian ‘subject’<sup>87</sup> to disciplinary power. Of course, we cannot overlook the fact that sedentary work of 14 h a day was performed by the prisoner,<sup>88</sup> and that the discipline of prisoner was done through the labour process. Moreover, we cannot forget that the panopticon was at the time seen to be a form of prison improvement which was more humane. It is also necessary to note that the principle of surveillance in the panopticon was not to ‘pry into the secret recesses of the heart’ of the prisoner but to confine ‘attention to *overt* acts, leav[ing] thoughts and fancies to their proper *ordinary*, the court *above*’.<sup>89</sup>

Yet because of transparent architectural layout of the panopticon, the disciplinary relations of ‘surveillant → prisoner’ and ‘surveillant → employee’ are not self-enclosed, but also visible to outside society.<sup>90</sup> As Foucault correctly pointed out, the panopticon was designed to be a ‘transparent building in which the exercise of power may be supervised by society as a whole.’<sup>91</sup> Bentham assumes that the surveyors are themselves objects of the surveillance by ‘the great *open committee* of the tribunal of the world’.<sup>92</sup> His ‘prison is transparent’.<sup>93</sup>

The surveillant who is put under the surveillance of ‘the tribunal of the world’<sup>94</sup> is required ‘to *disclose*, and even to print and *publish* his accounts of—the whole process

---

<sup>85</sup>Foucault pointed out that Bentham grafted the technical notion of an all-seeing gaze that exerts powers onto Rousseau’s dream of a transparent society. Foucault also compared Rousseau’s model (the watcher who makes friends) with that of Bentham’s model (the friend who becomes a watcher).

<sup>86</sup>Bentham (1838–43), IV, p. 44.

<sup>87</sup>Ichioka (1995), p. 69. Komatsu has criticized Foucault’s interpretation of the internalization of the eye and the code in panopticon, based on the contract system that operated within it. See Komatsu (2006).

<sup>88</sup>Bentham (1838–43), IV, p. 162.

<sup>89</sup>Bentham (1838–43), IV, p. 66.

<sup>90</sup>Bentham (1838–43), IV, p. 48.

<sup>91</sup>Foucault (1975), p. 209. Fukada has pointed out that the relation between the surveillant and the prisoner in Bentham’s Panopticon is replaced by the relation between the people and the civil servant in *Constitutional Code*. (Fukada 1984, p. 44).

<sup>92</sup>Bentham (1838–43), IV, p. 46.

<sup>93</sup>Bentham (1981), p. 229. Nakamura discusses the difference between Foucault of the twentieth century, for whom ‘surveillant delusion’ was the background for the focus on watched persons, and Bentham of the eighteenth century, for whom ‘voyeurism delusion’ was the background for focusing on the surveillant (Nakamura 1995).

<sup>94</sup>The public must not only observe the guards, but they must also bear the dual function of watching government officials and prisoners (Ishii 2002, p. 38). However, when prisoners were made a show of in public, Bentham proposed that their faces be covered with masks in consideration of their need to reintegrate into society (Bentham 1838–43, IV, p. 79). See Itai (2019).

and detail of his management—the whole history of the prison’.<sup>95</sup> If the surveillant violates this obligation of public openness, he will be penalized and his property seized. Bentham enumerates various strategies, such as contract management or single management, to prevent the surveillant’s arbitrary management of the prison.

### 3.4 *Surveillance and Publicity*

Let us now consider Bentham’s conception of government from the perspective of these external and heteronomous relations. First, the relation of surveillance and discipline to ‘the legislator →<sup>96</sup> the public’ naturally exists from the relation between the governors and the governed in society. The legislator as a governor has a role for curing people’s prejudices by political means.<sup>97</sup>

However, the public is not assumed to have a passive existence guided by the legislator alone.<sup>98</sup> For instance, an English member of parliament behaves as a good member because amity is cultivated with people overall and he interiorized their gaze.<sup>99</sup> Being in a relationship with a large number of people by the legislator creates consideration for them. Such a thing ‘is very useful for the reputation of the members of the assembly to be heard by impartial witnesses, and judged by a portion of the public which is changed every day’.<sup>100</sup> Moreover ‘the [legislator’s] internal censure will not be sufficient to secure probity, without the assistance of external censure’<sup>101</sup> by the public outside the assembly.

The public plays an active role as external disciplinary force upon the legislator: ‘the eye of the public makes the statesman virtuous’.<sup>102</sup> In Bentham’s condition of government, the public’s approval of government discussion is also the precondition of rule.<sup>103</sup> Bentham stated that ‘the stricter the dependence of the governors on the governed, the better will the government be, the larger and securer the measure of liberty’.<sup>104</sup> In fact, Bentham recognized that the trend of public opinion tends to correspond to the principle of utility in the age of enlightenment.<sup>105</sup>

---

<sup>95</sup>Bentham (1838–43), IV, p. 48.

<sup>96</sup>‘→’ indicates the direction of the disciplinary movement.

<sup>97</sup>Bentham (1996), p. 184.

<sup>98</sup>Nagai (1996), p. 186.

<sup>99</sup>Bentham (1996), p. 119. This is also a problem of the wish for approval (Lovejoy 1961). Regarding the ‘network of fame’ in Burke’s political thought, see Inuzuka (1997).

<sup>100</sup>Bentham (1999), p. 40. Bentham also stated that ‘the greater the communication among men, the greater the efficacy of the moral sanction’ (Bentham (1838–43), X, p. 145). Helvetius also values moral sanction. Morimura 1993, Chap. 3.

<sup>101</sup>Bentham (1999), p. 30.

<sup>102</sup>Bentham (1838–43), X, p. 145. See Peardon (1993), p. 640.

<sup>103</sup>Bentham (1999), p. 31.

<sup>104</sup>Bentham (2002), p. 415, Mack (1963), p. 455.

<sup>105</sup>Bentham (1996), pp. 35, 121, Bentham (1838–43), II, p. 424.

Thus, the relation of dependence and control between the governor and the governed is decidedly mutual. Therefore, the composition of the relation between surveillance and discipline in society is multilevel: ‘surveillant → prisoner’,<sup>106</sup> ‘surveillant → employee’, ‘public → surveillant’, ‘legislator → public’, and ‘public → legislator’.

The mutual relation of ‘legislator to public’ and ‘public to legislator’ is developed by the institution of publicity. First, to secure publicity is ‘to constrain the members of the assembly to perform their duty’ via the public tribunal, which is ‘more powerful than all the other tribunals’,<sup>107</sup> because the public have the power to recall legislators. This is a side of the surveillance and the discipline function that the non-substantial public tribunal demonstrates to the legislator. Next, publicity serves ‘to secure the confidence of the people, and their assent to the measures of the legislature’ which relates the security of society to the submission of the governed. Next is ‘to enable the governors to know the wishes of the governed’ and to enable the governed to know the action of the governors through publicity. Fourth, ‘in an assembly elected by the people, and renewed from time to time, publicity is absolutely necessary to enable the electors to act from knowledge’. Fifth, publicity serves ‘to provide the assembly with the means of profiting by the information of the public’. Sixth, publicity also serves ‘to reckon among the advantages of publicity, *the amusement which results from it*’.<sup>108</sup> In other words, the governor and the governed enact a controlled relation to each other through publicity.<sup>109</sup>

### 3.5 *The Enlightened Legislator and the Public*

The relations of ‘mutual enlightenment’ and ‘mutual discipline’ are chiefly caused by the character of the enlightened legislator and the public. The qualification for the legislator is to be an ‘enlightened man’.<sup>110</sup> Legislators are a small minority in any given society because human’s reason is not completely advanced.<sup>111</sup> ‘The end and the sole end which the legislator ought to have in view’, Bentham writes, is ‘the happiness of the individuals, of whom a community is composed, that is their pleasures and their security’.<sup>112</sup> To achieve this happiness, the legislator looks

---

<sup>106</sup>If we turn attention to the complaint system by which convict’s issued complaints to the surveillants, we see that the relation of ‘prisoner → surveillant’ existed here too. The watch principle of Bentham is not a one-sided, but a principle of mutual watch between the prisoner and the surveillant (Ishii 2002, p. 15).

<sup>107</sup>Bentham stated that ‘though this tribunal may err, it is incorruptible; that it continually tends to become enlightened’ (Bentham 1999, p. 29).

<sup>108</sup>Bentham (1999), pp. 29–34.

<sup>109</sup>Long (1977), pp. 198–199.

<sup>110</sup>Bentham (1931), p. 368.; Bentham (1802), iii, p. 19.

<sup>111</sup>Bentham (1970), p. 241.

<sup>112</sup>Bentham (1996), p. 34.

at society ‘with the eyes of statesmen’<sup>113</sup> and considers all members, augmenting happiness in the society. The legislator is an enlightened person who can apply the principle of utility, thus leading Bentham to assume that ‘the more men become enlightened, the more they will contract a spirit of general benevolence’.<sup>114</sup>

The act of interference allowed by the legislator is limited to the bounds of legislation. However, ‘the great difficulty here is, to persuade them to confine themselves within bounds. A thousand little passions and prejudices have led them to narrow the liberty of the subject in this line, in cases in which the punishment is either attended with no profit at all, or with none that will make up for the expense’.<sup>115</sup> The limitations of interference fall under the two areas of legislation and private ethics. The legislator is prohibited to interfere in the area of ‘private ethics’. If legislation exceeds the field and interferes in the matter of private ethics, a decrease in the overall happiness of the society will result. These occasions in which penalty exceeds the limits of interference are when it is ‘groundless’, ‘inefficacious’, ‘unprofitable’ and ‘needless’.<sup>116</sup> Legislation that falls under these categories is prohibited in terms of harm to society. Moreover, seen from another perspective, legislation is not an intervention from the point of ‘probity’ or when it is a precondition of ‘prudence’, though it is an intervention from the point of ‘beneficence’.<sup>117</sup> In spite of these restrictions, the legislator is forever prone to deviation.

Mischief also occurs when a man involved in government does not fulfil an obligatory act. Though all taxes are for securing order and the justice of a given nation as well as means ‘for the defence of the community against its foreign adversaries’,<sup>118</sup> a civil servant, military official or a judge can use this money corruptly and render its national function ineffective. Because of this, a supervisory organization to make the governor fulfil an obligatory act is sure to be needed. However, Bentham did not positively discuss this problem until his *Constitutional Code vol. I*, or *Official Aptitude Maximized; Expense Minimized* (1830), written in his later years.

In contrast, Bentham’s attitude towards the public is highly ambiguous. He pointed out the follies of the people<sup>119</sup> on the one hand and the wisdom of the people ‘in point of political discernment’ on the other.<sup>120</sup> This ambiguous attitude is derived from breaking the public down into three classes. ‘The first is composed of the most numerous party’ who ‘have not time to read, nor leisure for reasoning’, mainly because of the lack of money. ‘The second is composed of those who form a kind of judgment, but it is borrowed,’ namely who only follow someone’s opinion. The third is composed of those who judge for themselves, getting any information by

---

<sup>113</sup>Bentham (1996), p. 152.

<sup>114</sup>Bentham (1931), p. 432, Bentham (1802), iii, p. 129.

<sup>115</sup>Bentham (1996), p. 290.

<sup>116</sup>Bentham (1996), p. 159.

<sup>117</sup>Bentham (1996), pp. 283–293.

<sup>118</sup>Bentham (1996), p. 151.

<sup>119</sup>Bentham (1996), p. 119.

<sup>120</sup>Bentham (1977), p. 441.



themselves.<sup>121</sup> Thus, the reason for Bentham's ambiguous attitude towards the public is that he lumps these three classes together as 'the governed'.<sup>122</sup>

Bentham focused attention on the third class of public whose autonomous character makes them the most useful. He noted that 'the most useful plans have often been derived from private individuals'. Bentham had in his mind as figures such as 'Locke, Newton, Hume, [and] Adam Smith'.<sup>123</sup> He saw his role, as with these other private individuals, as that of a 'censor' who was to teach the science of legislation to the legislator.<sup>124</sup> By securing the representation of the public opinion through publication and the association, social discontent would be softened and the safety of public's submission and order enumerated.

### 3.6 *Rationality and Pessimism*

Having understood as delineated above the character of the legislator and the public, it is important that Bentham emphasized some measure of control for both. The legislator is required to push forward with legislation and stop his interference, while the public is asked to give their strict submission. The relation of mutual discipline of legislator and public results from these respects. It is necessary, then, to discuss Bentham's view of the human being in this context.

On the other hand, when the happiness of society is judged, the legislator who should follow only the principle of utility does not escape error. The felicific calculus as means of augmentation of social happiness cannot be strictly carried out for four reasons: a limit of human recognition ability,<sup>125</sup> a possibility of making a mistake of human calculation ability,<sup>126</sup> the gap between the felicific calculus performed prior to receiving pleasure and pain and the actual result,<sup>127</sup> and a difference in the sensibility of the individual.<sup>128</sup> As previously mentioned, the felicific calculus is insufficient in reasonable discussion by the legislator who might mistake this felicific calculus for some subjective interest.

Bentham, however, did not allow for the rule of irrationality in government, because he assumed that rational human beings chiefly participate in networks of 'mutual enlightenment' and 'mutual discipline.' Moreover, people with inferior

---

<sup>121</sup>Bentham (1999), p. 35. On the other hand, being enumerated as 'Enemy of publicity': 'the malefactor, who seeks to escape the notice of the judge; the tyrant, who seeks to stifle public opinion, whilst he fears to hear its voice; the timid or indolent man, who complains of the general incapacity in order to screen his own' (Bentham (1999), p. 30).

<sup>122</sup>About Bentham's distinguishing among three aptitudes that are desirable in voters, deputies, and officials, see Elster (2013), pp. 140–190.

<sup>123</sup>Bentham (1999), p. 33.

<sup>124</sup>Bentham (1977), pp. 397–8.

<sup>125</sup>Bentham (1996), p. 79; Bentham (1977), p. 492.

<sup>126</sup>Bentham (1931), p. 32; Bentham (1802), i, p. 53.

<sup>127</sup>Bentham (1996), pp. 40–1.

<sup>128</sup>Bentham (1996), p. 51.

reasoning capabilities do not play complete roles. In this sense, Bentham's trust in reason exists within the network of 'mutual enlightenment' and 'mutual discipline'. Sentiment is assumed to be conducted by reason.<sup>129</sup> Therefore, late Bentham divided public opinion tribunal into two sections: an 'aristocratic' section composed of better educated people and a 'democratic' section composed of those less educated.<sup>130</sup>

The network of 'mutual enlightenment' also has a deep relation to the procedure of felicific calculus as a means of the social reformation. The felicific calculus constantly involves the mechanism of 'feedback'<sup>131</sup> where further criticism and improvement by the governor and the governed functions ceaselessly under the institution of publicity.<sup>132</sup> This is because flexible correspondence is demanded of the legislator so that the judgment of right and wrong may change with the situation and the trends of the day.<sup>133</sup> The governor and the governed both participate in the procedure of felicific calculus through the network of 'mutual enlightenment'.

The relations between 'mutual enlightenment' and 'mutual discipline' thus result from the pessimistic element in Bentham's view on the human beings. These relations function through a publicity that secures safety and improvement of happiness in the society. The legislator and the public both form a controlled environment that leads the movement of human being's acts to 'certainty'.<sup>134</sup>

The descent process of social enlightenment is revealed in the metaphor 'by purifying the fountain, you will purify the streams'<sup>135</sup> (i.e., from the autonomous public class to other parts and from the assembly to various social clubs and meetings).<sup>136</sup> The ascent process of social enlightenment flows from the public to the assembly through publication, association, discussion, and petition. Such a dynamic government is supported by publicity. Bentham argued that 'without publicity, no good is permanent; under the auspices of publicity, no evil can continue',<sup>137</sup> and 'the grand security of securities is *publicity*'.<sup>138</sup> In such a way, 'every thing that is arbitrary in legislation vanishes',<sup>139</sup> and an increase of happiness in society becomes possible<sup>140</sup>.

---

<sup>129</sup>Viner (1958), p. 50.

<sup>130</sup>Bentham (1989), pp. 68–76.

<sup>131</sup>Nagai (1982), pp. 9–20.

<sup>132</sup>In Bentham's view, discussion held on the basis of the utility principle would become the common language of legislators and the citizens (Harrison 1983, p. 194). On the point of certainty of law as secured by publicity in Scottish judicial reform, see Draper (2004).

<sup>133</sup>Bentham (2002), p. 360.

<sup>134</sup>Ben-Dor interprets Bentham's constitutional theory as the 'Panoptic democracy' of a political system forever pressing for improvement. In this scenario, tribunals carried out by way of public opinion become a kind of surveillant (Ben-Dor (2000), p. 243).

<sup>135</sup>Bentham (1999), p. 35.

<sup>136</sup>Bentham (1999), p. 31.

<sup>137</sup>Bentham (1999), p. 37.

<sup>138</sup>Bentham (1838–43), II, p. 8.

<sup>139</sup>Bentham (1996), p. 274.

<sup>140</sup>See Kelly (1999), p. 95. See also Postema (1998), pp. 368–373.

The scheme in which the governor and the governed mutually control<sup>141</sup> and depend on one another was an inevitable result of Bentham's view on the human being. As Dinwiddy states it, 'Bentham did not—unlike Thompson and Marx—have any optimistic belief in the possibility of transforming' human conduct and nature.<sup>142</sup>

## 4 Conclusion—Pigeonhole of Selfishness

Bentham's utilitarianism allowed people to have a motivational drive to reflect on others, but since that drive was only 'a quiet murmur of morality', where political reflection (combining benefit and obligation)<sup>143</sup> was 'legal thunder,'<sup>144</sup> it was necessary to combine self-interest and duty through institutional design of legislation: 'if the thunders of the law probe impulse, the whispers of simple morality can have but little influence'.<sup>145</sup> Moreover, Bentham envisioned human behaviour that would incorporate panoptical architecture, 'indirect legislation' as a constraint on conduct selection, and oversight by and obedience to 'public opinion tribunal'.<sup>146</sup>

Indirect legislation relates to mutual relations of the governor and the governed. It comprises on the one hand the securing of safety and submission on the part of the governor, and on the other hand, the prevention of misrule on the part of the governed. The objects of indirect legislation are 'delinquency' and 'misrule'.<sup>147</sup> Early on, Bentham shifted priority from 'misrule' to 'delinquency', or from the governor's management to the governed's management, though Bentham approved of the influence of public opinion.

As a matter of fact, Bentham argued the problem of 'misrule' only in the 21st chapter, 'General Precautions against Abuses of Authority', of part IV, 'Indirect Means of Preventing Offences', in *The Theory of Legislation*, published in 1802. Therefore, although mutual relations between the governor and the governed are asymmetric, mutual and controlled relation between them do exist; the legislator

---

<sup>141</sup>Networks of discipline are not just limited to the governors or the public. Parents, especially the father, and teachers too serve to plant the idea of social order in children. See Bentham (1996), p. 68.

<sup>142</sup>Dinwiddy (1992), p. 434.; Roberts Jr. (1961), pp. 958–9.

<sup>143</sup>The principle of combining benefit and obligation is important in *Panopticon*. On this point, see Komatsu (2006).

<sup>144</sup>Bentham (1996), Chap. XVII, 11.

<sup>145</sup>Bentham (1996), Chap. XVII, 11. However, it is somewhat strange that in Bentham's case study, Coligny (Bentham 1996, Chap. X, 45) did not act under the influence of political sanctions.

<sup>146</sup>See Itai (2019).

<sup>147</sup>Bentham (1971), p. 127. From this we can conclude that for Bentham during this period, it is delinquency that is the primary concern of indirect legislation. Indeed, clear emphasis comes to be put on misrule in late Bentham. See Schofield (1996), pp. 223–234 and Itai (2002).

tries to secure the public's complete submission, and the public critically takes part in the prevention of misrule.<sup>148</sup>

But we have no difficulty in seeing that the governor is also the target of indirect legislation in Bentham's utilitarianism.<sup>149</sup> This shows that Bentham's notion of indirect legislation with respect to the conduct of both the ruled and the rulers fits into his conception of legislation in general as being directed towards the reduction of injury to public welfare.<sup>150</sup> Indeed, Bentham's famous Panopticon, with its design plan originally intended for a prison, was applied to the governmental buildings in his *Constitutional Code*.

While people are free to the extent that they do not harm others, delinquency and misrule, which impede the growth of social well-being, are controlled both through oversight and through physical mechanisms such as the Panopticon, legislators, and the general public.<sup>151</sup> In this view, the question of whether to regard people as selfishness or caring seems to recede into the background.

In fact, the motivation to take appropriate action is left to citizen's profit and loss accounting, or happiness calculation, and the maximization of happiness is aimed through representative democracy and other governance systems, without regard to the inner morality of citizens.<sup>152</sup> Bentham's utilitarianism was based on the idea that, in the process of freely pursuing their own interests, people should not infringe on the rights and freedoms of others, and that, as he stated in *Defence of a Maximum*<sup>153</sup> 'all eyes would be open to any violation of the law: all tongues ready to convey intelligence of it, would the law but be so adjusted as to give an compleat indemnity' they should participate proactively in the system of mutual surveillance of themselves in order to supervise violators in the market and in the city, thereby achieving appropriate governance aimed at the greatest happiness for the greatest number of people.<sup>154</sup>

The proposition that people's free actions result in social harmony, expressed in more detail, is from Smith's perspective, for instance, that people's free actions result in social benefits such as efficient production and fair prices through, for example, the protection of private property systems.<sup>155</sup> In Bentham's case, beside the protection of

---

<sup>148</sup>Ignatieff believes Bentham's plan for assembly reform and building the panopticon to be supplementary (Ignatieff 1989, p. 109). Komatsu points out that, 'It is exactly this idea that 'the principle of individualism is the starting point for the organization of collectivism' that is the shape of modern government' (Komatsu 2002).

<sup>149</sup>However, indirect legislation's relationship to the prevention of misrule is only touched upon in Chap. 19 of the Manuscripts and chapter 21 of Bentham (1802).

<sup>150</sup>See Bentham MS, lxxxvii, 2–4.

<sup>151</sup>See Itai (1998).

<sup>152</sup>See Ando (2010).

<sup>153</sup>Bentham (1952–54), iii, p. 259.

<sup>154</sup>This can be called 'liberalism without assuming personality' (Ando 2007).

<sup>155</sup>Schematically, Shaftesbury's argument that selfish behaviour is problematic because it produces undesirable results and Smith's argument that social behaviour is problematic because it produces undesirable results was there. Bentham attempted to demonstrate that both altruistic and selfish behaviour, based on desired utilitarian calculations, are mean to maximum happiness.

private property, the use of architectural affordances (the Panopticon) and networks of mutual surveillance is essential to the realization of the resulting social benefits.<sup>156</sup>

Bentham is not concerned about the problems of selfishness and order extensively investigated since Hobbes: whether society is an original contract or rooted in sympathy. Bentham presupposes the existing social order, and assumes that it does not matter whether human beings are selfish or altruistic, since, as Bentham himself acknowledged, a utilitarian system of governance transforms people into agents of appropriate utilitarian accounting and maximizes social well-being. Pleasure and pain are not predetermined, but their magnitude depends on their relation to the legal and social system. In this sense, Bentham may have changed the mainline tradition's way of considering selfishness.

## References

- Ando, K. 2007. *Governance and utility*, ed. Keiso Shobo (Japanese).
- Ando, K. 2010. Utilitarianism and Liberty, ed. Kitada, A., *Communication*, Iwanami Shoten (Japanese).
- Ben-Dor, O. 2000. *Constitutional limits and the public sphere: A critical study of Bentham's Constitutionalism*, (Intl Specialized Book Service).
- Bender, J. 1987. *Imagining the penitentiary: Fiction and the architecture of mind in eighteenth-century England*. University of Chicago Press.
- Bentham J. MS, Manuscripts in UCL.
- Bentham, J. 1977. *Fragment on government, in a comment on the commentaries and a fragment on government*, ed. J.H. Burns and H.L.A. Hart. The Athlone Press.
- Bentham, J. 1802. *Traité de législation civile et pénale*, 3 vols, ed. E. Dumont, Paris.
- Bentham, J. 1838–43. *The works of Jeremy Bentham*, ed. J. Bowring, Thoemmes.
- Bentham, J. 1931. *The theory of legislation*, ed. C.K. Ogden. 1931.
- Bentham, J. 1952–54. *Jeremy Bentham's economic writings*, 3 vols, ed. W. Stark, George Allen & Unwin Ltd.
- Bentham, J. 1970. *Of laws in general*, ed. H.L.A. Hart. The Athlone Press.
- Bentham, J. 1971. *Correspondence of Jeremy Bentham*, vol. iii, ed. I.R. Christie. The Athlone Press.
- Bentham, J. 1981. *Correspondence of Jeremy Bentham*, vol. iv, ed. A.T. Milne. The Athlone Press.
- Bentham, J. 1983. *Deontology together with a table of the springs of action and article on utilitarianism*, ed. A. Goldworth, Clarendon Press.
- Bentham, J. 1995. *Colonies, commerce, and constitutional law*. Clarendon Press.
- Bentham, J. 1996. IPML, *An introduction to the principles of morals and legislation*, ed. J.H. Burns and H.L.A. Hart. The Athlone Press, 1970; Paperback edition, Oxford, Clarendon Press.
- Bentham, J. 1998. *Legislator of the World: Writings on codification, law, and education*, ed. P. Schofield and J. Harris. Clarendon Press.
- Bentham, J. 1999. *Political tactics*, ed. C. Pease-Watkin, M. James, C. Blamires. Clarendon Press.
- Bentham. 2002. *Rights, representation, and reform: Nonsense upon stilts and other writings on the French revolution*, ed. P. Schofield, C. Pease-Watkin, C. Blamires. Clarendon Press.
- Burns, J.H. 1974. Bentham's critique of political fallacies. In *Jeremy Bentham critical assessments*, ed. B. Parekh, Routledge, 1993, vol. iii, 686–99.

---

<sup>156</sup>The type of surveillance used here is disciplined surveillance, which also involves physical transparency of buildings and supporting organizational operations.

- Crimmins, J. 1990. *Secular utilitarianism: Social science and the critique of religion in the thought of Jeremy Bentham*. Oxford University Press.
- Dinwiddy, J.R. 1982. Bentham on private ethics and the principle of utility. In *Jeremy Bentham Critical Assessments*, ed. B. Parekh, Routledge, 1993, vol. ii, 402–22.
- Dinwiddy, J.R. 1989. *Bentham*. Oxford University Press.
- Dinwiddy, J.R. 1992. *Radicalism and reform in Britain, 1780–1850*. The Hambledon Press.
- Draper, A.J. 2004. ‘Corruptions in the administration of justice: Bentham’s critique of civil procedure, 1806–1811. *Journal of Bentham Studies* 7.
- Elster, J. 1982. Sour grapes—utilitarianism and the genesis of wants. In *Utilitarianism and beyond*, ed. A. Sen and B. Williams. Cambridge University Press.
- Elster, J. 2013. *Securities against Misrule*. Cambridge University Press.
- Engelmann, S.G. 2003. *Imagining interest in political thought: Origins of economic rationality*. Duke University Press.
- Foucault, M. 1975. *Discipline and punish: The birth of the prison*, translated from the French by A. Sheridan, 2nd Vintage Books ed, 1995.
- Fukada, M. 1984. *Legal Positivism and utilitarianism*, Bokutakusha. (Japanese).
- Gaonkar, D.P., and R.J. McCarthy, Jr. 1994. Panopticism and Publicity Bentham’s Quest for Transparency. In *Public culture* vol. 6, 547–75.
- Habermas, J. 1990. *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*, Suhrkamp.
- Harrison, R. 1983. *Bentham*, Routledge & Kegan Paul.
- Ichioka, Y. 1995. Jeremy Bentham. In *A history of economic thought*, ed, Suzuki, Oota, Taka, and Yagi. Nagoya University Press (Japanese).
- Ignatieff, M. 1989 (1978). *A just measure of pain: The penitentiary in the Industrial Revolution, 1750–1850*, Penguin.
- Inuzuka, H. 1997. Edmund Burke, manners, and political power: politics of reputation, social relation, and refinement. *Journal of the Association of Political and Social Sciences*. 110(7–8): 607-67 (Japanese).
- Ishii, K. 2002. Treatment of Prisoners in Bentham’s Panopticon (1). *Ryukoku law Review*. 35(1) (Japanese).
- Itai, H. 1998. Conception of Government in Early Bentham: enlightened legislator and public. *Bulletin of Japanese Society for British Philosophy* 21 (Japanese).
- Itai, H. 2002. Hedonism and minority issue in Bentham’s Utilitarianism. *Annals of the Society for the History of Social Thought* 26 (Japanese).
- Itai, H. 2019. Surveillance and Metaphor of “Tribunal” in Bentham’s Utilitarianism. *Revue d’études benthamiennes* 16. <https://doi.org/10.4000/etudes-benthamiennes.6132>.
- Izumiya, S. 1976. J.S. Mill’s Introduction to the Study of ‘utilitarianism’, part 1, with a focus on the relationship between Bentham’s ethical beliefs. *Yokohama National University Journal of Humanities* 22.
- Kelly, P. 1999. Utilitarianism. In *Political concept context*, ed. Sato, S. and I. Soeya. Waseda University Press. (Japanese).
- Kodama, S. 1999. Virtue and happiness in Bentham’s Utilitarianism. *Studien zur Praktischen Philosophie* 22 (Japanese).
- Kodama, S. 2010. *Utility and intuition*. Keiso Shobo (Japanese).
- Komatsu, K. 2002. Reconsideration of Bentham’s Panopticon. *The journal of Ryūū Keizai University* 37(2): 19–29 (Japanese).
- Komatsu, K. 2006. *Governance and the education: Educational thought of Jeremy Bentham*. Ryutsu Keizai University Press (Japanese).
- Long, D. 1977. *Bentham on liberty: Jeremy Bentham’s idea of liberty in relation to his utilitarianism*. University of Toronto Press.
- Lovejoy, A.O. 1948. *Essays in the history of ideas*. The Johns Hopkins Press.
- Lyons, D. 1973. *In the interest of the government: A study of Bentham’s philosophy of utility and law*. Oxford University Press.

- Mack, M.P. 1962. *Bentham: an odyssey of ideas 1748–1792*. Heinemann Educational Books Ltd.
- Mill, J.S. 1838. Bentham. In *Jeremy Bentham critical assessments*, ed. B. Parekh, Routledge, 1993, vol. i, 142–75.
- Nagai, Y. 1982. *Bentham* (Man's intellectual heritage, 44), Kodan-sha (Japanese).
- Nagai, Y. 2003. *Bentham* (Critical biography of British intellectuals, 7), Kenkyu-sha (Japanese).
- Nakamura, H. 1996. Primal Scene of Panopticon. In *Komaba studies in society* 5: 49–68 (Japanese).
- Obata, S. 2015. Christianity and utilitarianism. In *The challenges of Jeremy Bentham*, ed. Fukagai, Y., and M. Kaino. Nakanishiya Publishing Co., Ltd. (Japanese).
- Parliamentary 1966. *Parliamentary history of England*, vol. xxix, 1791–1792. AMS Press.
- Peardon, T.P. 1951. Bentham's Ideal Republic. In *Jeremy Bentham critical assessments*, ed. B. Parekh, Routledge, 1993, vol. iii, 621–644.
- Plamenatz, J. 1949. *The English utilitarians*, Basil Blackwell.
- Postema, G.J. 1998. Bentham's equality-sensitive utilitarianism. *Utilitas* 10 (2): 144–158.
- Roberts, Jr. W. 1961. Bentham's conception of political change: A liberal approach. In *Jeremy Bentham Critical Assessments*, ed. B. Parekh, Routledge, 1993, vol. iii, 952–65.
- Rosen, F. 1992. *Bentham, Byron, and Greece, constitutionalism, nationalism, and early liberal political thought*. Clarendon Press.
- Rosen, F. 2003. *Classical utilitarianism from Hume to Mill*, Routledge.
- Schofield, P. 1996. Bentham on the identification of interest. *Utilitas* 8(2).
- Senga, S. 1988. Bentham's critique of Smith. In *The Scottish enlightenment*, ed. Tanaka S. Hokuju Shuppan (Japanese).
- Viner, J. 1958. Bentham and J.S. Mill: Utilitarian Background. In *Jeremy Bentham critical assessments*, ed. B. Parekh, Routledge, 1993, vol. iv.

# John Stuart Mill on Self-interest: Focusing on His Political Economy and the *Principles*



Yoshifumi Ozawa

**Abstract** This chapter reconstructs and provides an overview of John Stuart Mill's (1806–1873) ideas about self-interest, putting a special emphasis on what he called a political economy and on his *Principles of Political Economy*. Firstly, Mill explicitly threw light on the assumption of political economy. In his opinion, it presupposed that human beings always tried to obtain the greatest amount of wealth with as small a quantity of labor and abstinence as possible. Secondly, it was the desire for wealth rather than self-interest that Mill explicitly focused on when he defined political economy, although it was likely that Mill, in the *Principles*, mainly addressed people who endeavored to acquire wealth only for themselves. Thirdly, such selfish people did not reflect Mill's ideal and the third and subsequent editions of the *Principles* regarded the association of laborers among themselves as one of the means of cultivating those people's minds and provoking their public spirit in the future. Fourthly, in all the editions of the *Principles*, Mill, for the time being, endeavored to design political institutions which would reconcile the self-interested actions of individuals to the public benefit. This short-term and practical aspect of the *Principles*, though omitted in one of the abridged editions of it (i.e. Mill and Laughlin [1884] 1893), constituted Mill's art of government founded on his political economy.

**Keywords** John Stuart Mill · Self-interest · Public spirit · Political economy · Art of Government · Free trade · Education · Colonization · Association · James Laurence Laughlin

## 1 Introduction

The aim of this chapter is to reconstruct and provide an overview of John Stuart Mill's (1806–1873) ideas about self-interest, putting a special emphasis on what he called a political economy. It follows from what is shown in this chapter that these ideas in his *Principles of Political Economy* (seven editions: 1848, 1849, 1852, 1857, 1862, 1865

---

Y. Ozawa (✉)

Kyushu Sangyo University, 3-1 Matsukadai 2-Chome, Higashi-ku, Fukuoka, Japan  
e-mail: [yoshifumi.ozawa@gmail.com](mailto:yoshifumi.ozawa@gmail.com)



and 1871; henceforth *Principles*) were twofold. One related to the short term; Mill endeavored to design political institutions which would reconcile the self-interested actions of individuals to the public benefit. The other concerned the long term; Mill, especially in the third and subsequent editions of the *Principles*, attempted to cultivate individuals' minds and provoke their public spirit in the future. Of these two aspects, the latter makes a great difference between the eminent textbook on economics in the nineteenth century (i.e. Mill's *Principles*) and current famous textbooks on modern economics.

Section 2 focuses on the basis of Mill's opinions, namely his utilitarianism, and investigates his attitude toward selfishness and public spirit as well as his proposals for achieving the aim of utilitarianism. Section 3 clarifies Mill's definition of economics, or of what he called the political economy. Sections 4 and 5 address Mill's *Principles* (i.e. his most systematic treatise on the political economy); to explain further, Sect. 4 examines the short-term aspects of Mill's ideas about self-interest, whereas Sect. 5 tackles the long-term ones. There are some concluding remarks in Sect. 6.

## 2 Mill's Utilitarianism

### 2.1 Mill's Attitude Toward Selfishness and Public Spirit

The basis of Mill's opinions was his utilitarianism. To explain further, he endeavored to maximize the high-quality pleasure and minimize the pain of all concerned in the long run.<sup>1</sup> Mill emphasized that utilitarianism proposed to maximize the amount of happiness of *all concerned*. In other words, it assumed a "disinterested character" and aimed at "promoting the general interests of society" (Mill [1861b] 1969, 219). In his *Utilitarianism* (four editions: 1863, 1864, 1867 and 1871; revised and reprinted from Mill's articles published in 1861), Mill argued:

As between his own happiness [i.e. the agent's own happiness] and that of others, utilitarianism requires him to be as strictly impartial as a disinterested and benevolent spectator. [...] To do as one would be done by, and to love one's neighbour as oneself,<sup>2</sup> constitute the ideal perfection of utilitarian morality (ibid., 218; cf. Mill [1865] 1969, 335).

It may be worth pointing out that Mill presented a similar argument in the *Principles*: "In cases of actual or apprehended scarcity, many countries of Europe are accustomed to stop the exportation of food. [...] if the greatest amount of good to mankind on the whole, were the end aimed at in the maxims of international conduct, such collective churlishness would certainly be condemned by them" (Mill [1848]

---

<sup>1</sup>For further details of this point, see Mill's *System of Logic* (Mill [1843] 1974, 951), *On liberty* (Mill [1859] 1977, 224) and *Utilitarianism* (Mill [1861b] 1969).

<sup>2</sup>Mill's article of 1861 and the first edition of the *Utilitarianism* said "you," "your" and "yourself" instead of "one," "one's" and "oneself."

1965, 917). He went on to state that ideally “in such cases” people “should do to others what they would wish done to themselves” (*ibid.*, 918).<sup>3</sup>

Based on his utilitarianism, Mill placed a low value on the selfishness of mankind and a high one on humanity’s public-spirit. According to the *Utilitarianism*, “selfishness” meant a kind of human character which thought only about one’s own benefit, and Mill regarded it as one of the principal causes of unsatisfactory human lives:

When people who are tolerably fortunate in their outward lot do not find in life sufficient enjoyment to make it valuable to them, the cause generally is, caring for nobody but themselves. To those who have neither public nor private affections, the excitements of life are much curtailed, and in any case dwindle in value as the time approaches when all selfish interests must be terminated by death: while those who leave after them objects of personal affection, and especially those who have also cultivated a fellow-feeling with the collective interests of mankind, retain as lively an interest in life on the eve of death as in the vigour of youth and health (Mill [1861b] 1969, 215).

In the *Utilitarianism*, Mill looked down on selfish persons and said: “[...] no person of feeling and conscience would [consent to] be selfish and base” (*ibid.*, 211). His views on selfishness were also shown in other works such as his *Subjection of Women* (three editions: 1869, 1869 and 1870; henceforth *Subjection*) (e.g. Mill [1869] 1984, 288–289).

As for public spirit, Mill praised the execution of voluntary acts in the present state of society which increased the executor’s own pain (or at least did not increase his own pleasure), but which also increased the pleasure of others and, consequently, the amount of happiness in the world. Regarding “the morality of self-devotion,” Mill argued in the *Utilitarianism*: “The only self-renunciation which” the utilitarian morality “applauds, is devotion to the happiness, or to some of the means of happiness, of others” (Mill [1861b] 1969, 218). According to the *Utilitarianism*, “[t]hough it is only in a very imperfect state of the world’s arrangements that any one can best serve the happiness of others by the absolute sacrifice of his own, yet so long as the world is in that imperfect state, I fully acknowledge that the readiness to make such a sacrifice is the highest virtue which can be found in man” (*ibid.*, 217). In this context, it must be emphasized that voluntariness or spontaneity was a necessary condition and that Mill did not regard the above-mentioned self-sacrifice as a duty. According to Mill’s *Auguste Comte and Positivism* (two editions: 1865 and 1866; revised and reprinted from Mill’s articles published in 1865; henceforth *Positivism*), Auguste Comte (1798–1857) coined the term “altruism” and argued: “We should endeavour not to love ourselves at all” (Mill [1865] 1969, 335). In other words, Comte’s golden rule of morality was not “[t]o do as we would be done by, and to love our neighbour as ourself,” but “to live for others” (*ibid.*, 335). Mill disagreed with this point of view and focused on “the region of positive worthiness” which was intermediate between “the region of duty and that of sin” (*ibid.*, 337):

---

<sup>3</sup>Focusing on the *Principles*, Sugihara (1985; 2003, 311–335) and Mawatari (1997a, 401–402, 408, 415) regard Mill’s views on infant industry protection and on colonization as his attempts to maximize the amount of good to mankind.

There is a standard of altruism to which all should be required to come up, and a degree beyond it which is not obligatory, but meritorious. It is incumbent on every one to restrain the pursuit of his personal objects within the limits consistent with the essential interests of others. [...] If in addition to fulfilling this obligation, persons make the good of others a direct object of disinterested exertions, postponing or sacrificing to it even innocent personal indulgences, they deserve gratitude and honour, and are fit objects of moral praise (ibid., 337–338).

## 2.2 *Means of Maximizing the Amount of Happiness in the World*

What the preceding section (Sect. 2.1) immediately makes clear is that it was of vital importance in Mill's opinion to cultivate human minds and turn selfish persons into public-spirited ones as far as possible; the best account of this point can be found in the following quotation from the *Utilitarianism*: "Utilitarianism [...] could only attain its end by the general cultivation of nobleness of character" (Mill [1861b] 1969, 213–214). In fact, Mill proposed to take two measures in order to maximize the happiness of all concerned. One of them was education in a broad sense, namely the cultivation of human minds, the other being political reform. After pointing out that "the golden rule of Jesus of Nazareth" constituted "the ideal perfection of utilitarian morality," Mill went on to say:

As the means of making the nearest approach to this ideal, utility would enjoin, first, that laws and social arrangements should place the happiness, or [...] the interest, of every individual, as nearly as possible in harmony with the interest of the whole; and secondly, that education and opinion, which have so vast a power over human character, should so use that power as to establish in the mind of every individual an indissoluble association between his own happiness and the good of the whole; especially between his own happiness and the practice of such modes of conduct, negative and positive, as regard for the universal happiness prescribes: so that not only he may be unable to conceive the possibility of happiness to himself, consistently with conduct opposed to the general good, but also that a direct impulse to promote the general good may be in every individual one of the habitual motives of action, and the sentiments connected therewith may fill a large and prominent place in every human being's sentient existence (ibid., 218; cf. ibid., 231–232).

"A direct impulse to promote the general good" and "the sentiments connected therewith" can be called the public spirit.

Firstly, it may be worth mentioning that Mill referred to the possibility of turning selfish persons into public-spirited ones as well as the difficulty of maintaining such spirit for life. The former was expressed best by Mill when he, in his posthumous *Autobiography* (1873), remarked as follows:

Interest in the common good is at present so weak a motive in the generality, not because it can never be otherwise, but because the mind is not accustomed to dwell on it as it dwells from morning till night on things which tend only to personal advantage. When called into activity as only self interest now is, by the daily course of life, and spurred from behind by the love of distinction and the fear of shame, it is capable of producing, even in common

men, the most strenuous exertions as well as the most heroic sacrifices (Mill [1873] 1981, 241).

In the *Utilitarianism*, Mill also denied “an inherent necessity that any human being should be a selfish egotist,” and went on to express: “Genuine private affections, and a sincere interest in the public good, are possible, though in unequal degrees, to every rightly brought up human being” (Mill [1861b] 1969, 216). Added to these, the third and subsequent editions of the *Principles* argued: “Mankind are capable of a far greater amount of public spirit than the present age is accustomed to suppose possible” (Mill [1848] 1965, 205). In order to cultivate such a public spirit in people, Mill attached great importance to the role of education: “Education, habit, and the cultivation of the sentiments will make a common man dig or weave for his country, as readily as fight for his country” (Mill [1873] 1981, 239). The following quotation from Mill’s *Autobiography* also indicated his opinion that selfish people could be made into public-spirited ones through education: “[...] so long as education continues to be so wretchedly imperfect, we dreaded [...] the selfishness and brutality of the mass” (ibid., 239). Additionally, Mill remarked in the *Positivism*:

As a rule of conduct, to be enforced by moral sanctions, we think no more should be attempted than to prevent people from doing harm to others, or omitting to do such good as they have undertaken. [...] But above this standard there is an unlimited range of moral worth, up to the most exalted heroism, which should be fostered by every positive encouragement, though not converted into an obligation. It is as much a part of our scheme as of M. Comte’s, that the direct cultivation of altruism, and the subordination of egoism to it, far beyond the point of absolute moral duty, should be one of the chief aims of education, both individual and collective (Mill [1865] 1969, 339).<sup>4</sup>

With regard to the difficulty mentioned above, Mill, in the *Utilitarianism*, pointed out: “Capacity for the nobler feelings is in most natures a very tender plant, easily killed, not only by hostile influences, but by mere want of sustenance” (Mill [1861b] 1969, 213).

What is more, it is vital to bear in mind that the political reform which would bridge the gap between the selfish behaviors of individuals and the public interest was no less essential in Mill’s views than education. According to the *Principles*, human character could be changed very gradually: “no great change in the habits of mankind is otherwise than difficult, and in general either painful or very slow” (Mill [1848] 1965, 12). In the *Autobiography*, Mill also asserted: “it is only by slow degrees, and a system of culture prolonged through successive generations, that men in general can be brought up” to work hard for their country (Mill [1873] 1981, 239–241). Mill subsequently revealed that new social institutions based on the public spirit of all people were not realistic for the time being. In other words, he did not “overlook the folly of premature attempts to dispense with the inducements of private

---

<sup>4</sup>In this context, Mill referred to the merit of military training which every citizen received in antiquity (Mill [1865] 1969, 339; cf. Mill [1848] 1965, 102–103). For Mill’s views on military training, see also Mill ([1871] 1988) and Varouxakis (2013a, 135–136; 2013b, 151–153, 164–171, 182–183).

interest in social affairs, while no substitute for them has been or can be provided” (ibid., 241).

One final point is that Mill, in addition to his education and political reforms, regarded the eradication of poverty as one of the means of increasing the amount of happiness in the world. According to the *Utilitarianism*, “utility includes not solely the pursuit of happiness, but the prevention or mitigation of unhappiness” (Mill [1861b] 1969, 214). Mill regarded it as necessary to contest with “the great sources of physical and mental suffering—such as indigence [...]” (ibid., 216). In Mill’s opinion, “[p]overty [...] may be completely extinguished by the wisdom of society, combined with the good sense and providence of individuals” (ibid., 216). It should also be added that the eradication of poverty is, according to the *Principles*, a necessary condition for education to prosper: “Education is not compatible with extreme poverty” (Mill [1848] 1965, 375).

With regard to education, political reforms and the eradication of poverty, we proceed to investigate Mill’s opinions, focusing on his political economy and *Principles*.

### 3 Mill’s Definition of a Political Economy

Mill’s most systematic treatise on political economy was *Principles*.<sup>5</sup> According to this voluminous work, political economy addressed “the nature of Wealth, and the laws of its production and distribution” (Mill [1848] 1965, 3). In his *Inaugural Address Delivered to the University of St. Andrews* (two editions: 1867 and 1867; henceforth *Inaugural Address*), Mill also argued that political economy had been sufficiently established as a branch of science and that it could play a vital role in political reforms (Mill [1867] 1984, 245).

In addition to these, Mill, as is well known, defined political economy more clearly in his *Essays on Some Unsettled Questions of Political Economy* (1844; henceforth *Essays*), particularly in its fifth article (a reprint, with some alterations, from his article published in 1836); and it should be noted that in all the editions of his *System of Logic* (eight editions: 1843, 1846, 1851, 1856, 1862, 1865, 1868 and 1872; henceforth *Logic*), Mill quoted from his article of 1836 or the fifth one of the *Essays*. Mill’s definition of political economy established: “The science which traces the laws of such of the phenomena of society as arise from the combined operations of mankind for the production of wealth, in so far as those phenomena are not modified by the pursuit of any other object” (Mill [1844] 1967, 323).<sup>6</sup> In Mill’s

<sup>5</sup>Mill generally called what is now called economics political economy. In the *Principles*, however, he used the word “pure economics” just once (Mill [1848] 1965, 327).

<sup>6</sup>After pointing out this definition, Mill continued: “[...] the didactic writer on the subject will naturally combine in his exposition, with the truths of the pure science, as many of the practical modifications as will, in his estimation, be most conducive to the usefulness of his work” (Mill [1844] 1967, 323). It is, therefore, reasonable that Mill’s *Principles* included not only the truths of political economy as a pure science but also their practical modifications. In fact, Mill, in his

opinion, political economy was “essentially an *abstract* science” and presupposed “an arbitrary definition of man, as a being who invariably does that by which he may obtain the greatest amount of necessities, conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with which they can be obtained in the existing state of knowledge” (ibid., 325–326). Mill’s political economy only took “the desire of wealth” and “two perpetual counter-motives” into consideration, the latter being an “aversion to labour” and the “desire of the present enjoyment of costly indulgences” (ibid., 321–322; quoted in Mill [1843] 1974, 902).

Mill’s *Principles* and the fifth article of his *Essays* provide some additional explanation of the above-mentioned definition. Firstly, the word “wealth” was defined by Mill in *Principles* as “all useful or agreeable things except those which can be obtained, in the quantity desired, without labour or sacrifice” (Mill [1848] 1965, 10). Added to this, according to *Principles*, “what are called immaterial products” should not be regarded as wealth (ibid., 10, 49).

Secondly, Mill viewed political economy not as an art but a science. With regard to the distinction between science and art, Mill, in the *Essays*, explained as follows:

The one deals in facts, the other in precepts. Science is a collection of *truths*; art, a body of *rules*, or directions for conduct. The language of science is, This is, or, This is not; This does, or does not, happen. The language of art is, Do this; Avoid that. Science takes cognizance of a *phenomenon*, and endeavours to discover its *law*; art proposes to itself an *end*, and looks out for *means* to effect it (Mill [1844] 1967, 312; cf. Mill [1843] 1974, 943–946, 949–950).

Additionally, the *Essays* argued that art was based on a variety of sciences: “the necessary foundation of all art is science” and “[e]ach art presupposes, not one science, but science in general; or, at least, many distinct sciences” (Mill [1844] 1967, 331; cf. Mill [1843] 1974, 947–948).

Thirdly, according to the *Essays*, political economy formed a part of speculative politics and so constituted one of the foundations of practical politics or the art of government. Mill argued that science was divided into physical science and mental, moral or psychological one; the latter treating of “the laws of mind, and of all complex phenomena in so far as dependent upon the laws of mind” (Mill [1844] 1967, 317). The latter science comprised three parts and one of them examined “[those] principles of human nature which are peculiarly connected with the ideas and feelings generated in man by living in a state of *society*, that is, by forming part of a union or aggregation of human beings for a common purpose or purposes.” Mill labeled this branch of science as follows: (1) “*social economy*,” (2) “*speculative politics*, or the *science* of politics, as contradistinguished from the art,” (3) “the natural history of society,” (4) “the social science” and (5) “social philosophy” (ibid., 320, 333–335). To explain further, such science was called “speculative politics, as being the scientific foundation of practical politics, or the art of government, of which the art of legislation is a part” (ibid., 320–321).<sup>7</sup> With regard to political economy, the *Essays*

---

*Autobiography*, remarked that *Principles* “was not a book merely of abstract science, but also of application” (Mill [1873] 1981, 243).

<sup>7</sup>Mill also used the word “the art of government” in his *On Liberty* (Mill [1859] 1977, 308–309) and *Considerations on Representative Government* (Mill [1861a] 1977, 393). With regard to the word

argued: “What is now commonly understood by the term “Political Economy” is not the science of speculative politics, but a branch of that science. [...] It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth” (ibid., 321).

Fourthly, Mill and other political economists recognized that the aforementioned supposition of political economy failed to correspond with real human beings, and in practice they thought it necessary to modify the conclusions of political economy by considering motives other than the desire of wealth. According to the *Essays*, no political economist ever imagined “that real men had no object of desire but wealth” (Mill [1844] 1967, 327). In Mill’s opinion, such supposition was a part of “the mode in which science must necessarily proceed.” He went on to say: “So far as it is known, or may be presumed, that the conduct of mankind in the pursuit of wealth is under the collateral influence of any other of the properties of our nature than the desire of obtaining the greatest quantity of wealth with the least labour and self-denial, the conclusions of Political Economy will so far fail of being applicable to the explanation or prediction of real events, until they are modified by a correct allowance for the degree of influence exercised by the other cause” (ibid., 322, 323; quoted in Mill [1843] 1974, 902, 903).

Fifthly, although Mill, for the time being, placed a high value on many people’s desire for wealth, his ideal state was the stationary one in which people no longer wished to acquire more wealth. It is likely, therefore, that in Mill’s opinion, the supposition of political economy did not correspond with ideal human beings and political economy played a central role in political reforms only in the current state of society, in which almost all people struggled to increase their wealth. Indeed, Mill said in *Principles*: “That the energies of mankind should be kept in employment by the struggle for riches [...] until the better minds succeed in educating the others into better things, is undoubtedly more desirable than that they should rust and stagnate” (Mill [1848] 1965, 754). However, *Principles* referred to “the deeper consideration [...] that the more wholesome state of society is [...] that in which the greatest possible numbers possess and are contented with a moderate competency, which all may hope to acquire” (ibid., 891). In the famous chapter of *Principles*, Mill clearly said: “I sincerely hope, for the sake of posterity, that they [i.e. wealth and population] will be content to be stationary, long before necessity compels them to it” (ibid., 756). According to Mill, it is ultimately desirable “to devote to labour, for mere pecuniary gain, fewer hours in the day, fewer days in the year, and fewer years of life” (ibid., 106); and in England, therefore, “it is not the desire of wealth that needs to be taught, but the use of wealth, and appreciation of the objects of desire which wealth cannot purchase” (ibid., 105). As for England, Mill also remarked: “though the general thriftiness of the labouring class is much below what is desirable, the spirit

---

“practical politics,” Mill argued in the *Logic*: “The aim of practical politics is to surround any given society with the greatest possible number of circumstances of which the tendencies are beneficial, and to remove or counteract, as far as practicable, those of which the tendencies are injurious” (Mill [1843] 1974, 898).



of accumulation in the more prosperous part of the community requires abatement rather than increase” (ibid., 187).

Finally, Mill’s political economy did not necessarily presuppose selfish people, namely those who only took account of their own wealth. In *Principles*, he referred to not only selfish motives but also altruistic ones for abstinence and accumulation:

The savings by which an addition is made to the national capital, usually emanate from *the desire of persons* to improve what is termed their condition in life, or *to make a provision for children or others, independent of their exertions*. [...] there is in every country some rate of profit, below which persons in general will not find *sufficient motive to save for the mere purpose* of growing richer, or *of leaving others better off than themselves* (emphasis added; Mill [1848] 1965, 737).

Mill also remarked that not a little saving was caused in practice by the altruistic motives:

Deficient strength of the desire of accumulation may arise from improvidence, or from want of interest in others. [...] The effect of want of interest in others in diminishing accumulation will be admitted, if we considered how much saving at present takes place, which has for its object the interest of others rather than of ourselves; the education of children, their advancement in life, the future interests of other personal connexions, the power<sup>8</sup> of promoting, by the bestowal of money or time, objects of public or private usefulness (ibid., 163–164).

Added to these, Mill gave an example of the production of wealth for the use of others, focusing on “the Indians of Paraguay” ruled by “the Jesuits” (ibid., 166, 210–211). However, it is reasonable to suppose that in the *Principles* Mill largely analyzed selfish people who endeavored to acquire wealth for themselves, because he pointed out that “the selfish type of character” was “formed by the present standard of morality, and fostered by the existing social institutions” (ibid., 210).<sup>9</sup> It should also be added that, according to the *Principles*, “the most numerous class” was “the class of manual labourers” (Mill [1848] 1965, 758; cf. ibid., 357) and that unskilled laborers did not have the sufficient motive for saving for the sake of their children’s education: “In England, and most European countries, elementary instruction cannot be paid for, at its full cost, from the common wages of unskilled labour, and *would not if it could*” (emphasis added; ibid., 949).

---

<sup>8</sup>The fifth and previous editions of the *Principles* said “desire” instead of “power.”

<sup>9</sup>In his *Autobiography*, Mill also remarked: “The deep rooted selfishness [...] forms the general character of the existing state of society” (Mill [1873] 1981, 241).



## 4 The Short-Term Aspect of the *Principles*: The Distinction Between *Laissez-Faire* and Government Intervention

### 4.1 The General Rule

In the eleventh chapter of the fifth book of the *Principles*, Mill endeavored to make a distinction between functions which governments ought to perform and those which they ought not. Mill, in the *Principles*, argued that “[o]ne of the most disputed questions both in political science and in practical statesmanship” was “to what departments of human affairs” governments’ authority “should extend” (Mill [1848] 1965, 799). That chapter of the *Principles* focused on this question and investigated “the line of demarcation between the things with which government should, and those with which they should not, directly interfere,” “by examining [...] what are the advantages, and what the evils or inconveniences, of government interference” (ibid., 880, 937). The contents of that chapter of the *Principles* were also illustrated in Mill’s *Considerations on Representative Government* (three editions: 1861, 1861 and 1865; henceforth *Considerations*). He mentioned: “I have said elsewhere what seemed to me most essential respecting the principles by which the extent of” governmental action “ought to be determined” (Mill [1861a] 1977, 534). The word “elsewhere” meant “*On Liberty*, concluding chapter; and, at greater length, in the final chapter of *Principles of Political Economy*” (ibid., 534), that is to say the fifth chapter of Mill’s *On liberty* (four editions: 1859, 1859, 1864 and 1869; henceforth *Liberty*) and the eleventh chapter of the fifth book of the *Principles*.

In the first part of “the final chapter of *Principles of Political Economy*,” Mill adduced five “evils or inconveniences of government interference” (Mill [1848] 1965, 937). Based on these five “[o]bjections to government intervention,” he regarded “[*laissez-faire*” as “the general rule” (ibid., 937, 944):

The preceding are the principal reasons, of a general character, in favour of restricting to the narrowest compass the intervention of a public authority in the business of the community: and few will dispute the more than sufficiency of these reasons, to throw, in every instance, the burthen of making out a strong case, not on those who resist, but on those who recommend, government interference. *Laissez-faire*, in short, should be the general practice: every departure from it, unless required by some great good, is a certain evil (ibid., 944–945).

One of those five “reasons of the non-interference principle” was that “in all the more advanced communities, the great majority of things are worse done by the intervention of government, than the individuals most interested in the matter would do them, or cause them to be done, if left to themselves” (Mill [1848] 1965, 941, 948). Broadly speaking, the grounds for this reason was that “people understand their own business and their own interests better, and care for them more, than the government does, or can be expected to do” (ibid., 942). Furthermore, it may be worth mentioning that Mill also presented a similar argument in his *Subjection*: “The modern conviction [...] is, that things in which the individual is the person directly

interested, never go right but as they are left to his own discretion [...]” (Mill [1869] 1984, 273).

Mill argued that the reason behind the non-interference principle mentioned above was closely related to political economy. In the *Principles*, he clarified that “what was represented as the great principle of political economy” was “that individuals are the best judges of their own interest” (Mill [1848] 1965, 959; cf. *ibid.*, 950–951, 953). Added to this, Mill, in the *Liberty*, referred to cases “when the thing to be done is likely to be better done by individuals than by the government;” and he said that these cases had been “sufficiently enlarged upon by political economists” (Mill [1859] 1977, 305). According to the *Considerations*, “if the principles of modern politics and political economy are good for anything, it is for proving that these points [i.e. what individuals are and are not fit for, and what they shall and shall not be allowed to attempt] can only be rightly judged of by the individuals themselves [...]” (Mill [1861a] 1977, 479). In the *Subjection*, Mill showed that the “general principle of social and economical science” was that “individuals, with such help as they can derive from the opinion of those who know them, are [...] better judges than the law and the government, of their own capacities and vocation” (Mill [1869] 1984, 274).

Related to the proposition “that each is the only safe guardian of his own rights and interests,” Mill, in the *Considerations*, remarked that such a proposition never indicated that selfishness is humanity’s universal nature, and insisted that it was not “the doctrine of the general predominance of self-interest” (Mill [1861a] 1977, 404–405). On the proposition discussed here, he declared:

Many, indeed, have a great dislike to it as a political doctrine, and are fond of holding it up to obloquy, as a doctrine of universal selfishness. To which we may answer, that whenever it ceases to be true that mankind, as a rule, prefer themselves to others, and those nearest to them to those more remote, from that moment Communism is not only practicable, but the only defensible form of society; and will, when that time arrives, be assuredly carried into effect. For my own part, not believing in universal selfishness, I have no difficulty in admitting that Communism would even now be practicable among the *élite* of mankind, and may become so among the rest (*ibid.*, 404–405).<sup>10</sup>

According to the *Inaugural Address*, therefore, political economy could help us to anticipate beneficial consequences from humanity’s self-interest but would not make public-spirited people selfish: “My advice to you is to study the great writers on Political Economy, and hold firmly by whatever in them you find true; and depend upon it that if you are not selfish or hard-hearted already, Political Economy will not make you so” (Mill [1867] 1984, 245).<sup>11</sup>

---

<sup>10</sup> According to the *Principles*, “the Communistic doctrine [...] forms the extreme limit of Socialism; according to which not only the instruments of production [...] are the joint property of the community, but the produce is divided and the labour apportioned, as far as possible, equally” (Mill [1848] 1965, 210). For further details of Mill’s views on Communism, see the third and subsequent editions of the *Principles* (*ibid.*, xciii, 199–209).

<sup>11</sup> This passage is quoted at the beginning of one of the present famous textbooks on modern economics (i.e. Mankiw [1991] 2019, vii).

Focusing on political economy, let us devote a little more space to examining the “superior efficiency of private agency, owing to stronger interest in the work” (Mill [1848] 1965, 941). As has been noted, Mill said that the “concluding chapter” of the *Liberty* discussed essential points on the principles which ought to determine the extent of governmental action (Mill [1861a] 1977, 534). It is, therefore, useful to quote from that chapter:

Speaking generally, there is no one so fit to conduct any business, or to determine how or by whom it shall be conducted, as those who are personally interested in it. This principle condemns the interferences, once so common, of the legislature, or the officers of government, with the ordinary processes of industry (Mill [1859] 1977, 305).

“[T]he doctrine of Free Trade” provides an example. According to the *Liberty*, this doctrine had been brought forth by political economists and argued “that both the cheapness and the good quality of commodities are most effectually provided for by leaving the producers and sellers perfectly free, under the sole check of equal freedom to the buyers for supplying themselves elsewhere.” Mill went on to state that although “it was once held to be the duty of governments [...] to fix prices, and regulate the processes of manufacture,” such a doctrine now makes it clear that these restraints “are wrong [...] because they do not really produce the results which it is desired to produce by them” (ibid., 293). Furthermore, Mill remarked not only in the *Principles* but also in the *Subjection* that on the grounds of the doctrine of free trade, those restraints on industrial operations had been abolished “in even the least enlightened country of the European commonwealth of nations (Mill [1848] 1965, 946):

Among the industrious classes, [...] nobody could practise any calling deemed important, in any but the legal manner [...]. In modern Europe, [...] diametrically opposite doctrines now prevail. Law and government do not undertake to prescribe by whom any social or industrial operation shall or shall not be conducted, or what modes of conducting them shall be lawful. These things are left to the unfettered choice of individuals (Mill [1869] 1984, 273; cf. Mill [1848] 1965, 926, 944–946).

In the *Principles*, for example, Mill justified the business of self-interested speculators such as corn-dealers and generally advocated free trade for the public benefit. To explain further, the business of those speculators was “to buy goods in order to resell them at a profit” and made it possible for the surplus goods of one year to remain “to supply the deficiency of another” (Mill [1848] 1965, 715). According to Mill, “[t]he interest [...] of the speculators as a body, coincides with the interest of the public; and as they can only fail to serve the public interest in proportion as they miss their own, the best way to promote the one is to leave them to pursue the other in perfect freedom” (ibid., 717; cf. ibid., 926–927).

As for foreign trade, Mill, in the *Essays*, regarded “the doctrine of Freedom of Trade between nation and nation” as “one of the most valuable results of modern political philosophy” (Mill [1844] 1967, 231). According to the *Principles*, the most notable of “[t]he false theories of political economy” which had caused “so much mischief in times past” was “the doctrine of Protection to Native Industry.” Based on

this doctrine, foreign free trade was criticized and “the prohibition, or the discouragement by heavy duties, of such foreign commodities as” were “capable of being produced at home” was advocated. In short, it was clarified that “the interest of the consumer appeared [...] to be contrary to the public interest” (Mill [1848] 1965, 913–914). In the *Principles*, however, Mill investigated “the effects of international trade” and identified that such a protective doctrine was false: “[...] the importation of foreign commodities, in the common course of traffic, never takes place, except when it is, economically speaking, a national good, by causing the same amount of commodities to be obtained at a smaller cost of labour and capital to the country.” Mill, therefore, opposed the prohibition and discouragement of the foreign trade mentioned above: “To prohibit” the importation of foreign commodities, “or impose duties which prevent it, is to [...] compel a waste, of the difference between the labour and capital necessary for the home production of the commodity, and that which is required for producing the things with which it can be purchased from abroad” (ibid., 914; cf. ibid., 591, 922). Additionally, Mill pointed out: “Commerce is virtually a mode of cheapening production; and in all such cases the consumer is the person ultimately benefited” (ibid., 593). To summarize, Mill insisted that “the interest of the consumer” would be consistent with “the public interest” (ibid., 914).<sup>12</sup>

## 4.2 Large Exceptions to Laissez-Faire

After revealing five objections to government intervention, Mill turned to the second part of his task. In the last part of the *Principles*, Mill showed “[l]arge exceptions to *laissez-faire*” in which government intervention would be recommended (ibid., 947). To explain more comprehensively, Mill mainly addressed those cases to which one of the objections did not apply; this objection was that which has been focused on in Sect. 4.1 above (Mill [1848] 1965, 946–47, 950–51). In such cases, things would not be better done by the individuals most interested in the matter than by a government. To provide an outline, Mill, in the *Principles*, focused on “the exceptions to the practical maxim, that the business of society can be best performed by private and voluntary agency” (ibid., 970).

For the moment, let us closely examine two cases: that of education and that of colonization. According to the *Principles*, both cases belonged to the exceptions to *laissez-faire* mentioned above. Additionally, with regard to Mill’s views on the eradication of poverty, those two cases deserve investigation. For the purpose of eradicating poverty, Mill attached great importance to laborers’ voluntary restriction of the increase of their numbers (e.g. Mill [1873] 1981, 107–109); in order to realize such restrictions, Mill, in the *Principles*, proposed “[t]wofold means of elevating

---

<sup>12</sup>In the *Considerations*, Mill suggested that England had “completely outgrown” “mutual exclusion by hostile tariffs” (Mill [1861a] 1977, 565). According to the *Autobiography*, “the last vestiges” of the protective system “were [...] swept away by Mr. Gladstone in 1860” (Mill [1873] 1981, 103). This fiscal reform by Gladstone was reflected in the fifth and subsequent editions of the *Principles* (Mill [1848] 1965, 871–872).

the habits of the labouring people,” namely education and colonization (Mill [1848] 1965, 374–379). On this point, he remarked as follows:

For the purpose [...] of altering the habits of the labouring people, there is need of a twofold action, directed simultaneously upon their intelligence and their poverty. An effective national education of the children of the labouring class, is the first thing needful: and, coincidentally with this, a system of measures which shall [...] extinguish extreme poverty for one whole generation (ibid., 374).

In this context, the word “education” meant the cultivation of “common sense” rather than that of public spirit: “[...] the aim of all intellectual training for the mass of the people, should be [...] to qualify them for forming a sound practical judgment of the circumstances by which they are surrounded” (ibid., 375).

As for the case of education, Mill argued that a government ought to assist in the education of the masses largely because the uneducated could not understand the necessity for nor the value of education. In other words, Mill regarded such a case as one of the exceptions to “the proposition that the consumer is a competent judge of the commodity” and, therefore, as “one of those things which it is admissible in principle that a government should provide for the people” (Mill [1848] 1965, 947, 948). According to the *Principles*, “[t]hose who most need to be made wiser and better, usually desire it least, and if they desired it, would be incapable of finding the way to it by their own lights” (ibid., 947). On the other hand, Mill had a certain confidence in government-provided education for uneducated citizens: “any well-intentioned and tolerably civilized government may think, without presumption, that it does or ought to possess a degree of cultivation above the average of the community which it rules, and that it should therefore be capable of offering better education and better instruction to the people, than the greater number of them would spontaneously demand<sup>13</sup>” (ibid., 947–948). However, it should be noted that Mill opposed a state monopoly of education: “Though a government [...] may, and in many cases ought to, establish schools and colleges, it must neither compel nor bribe any person to come to them; nor ought the power of individuals to set up rival establishments, to depend in any degree upon its authorization” (ibid., 950).

Regarding the case of colonization, Mill insisted that a government ought to intervene, especially in the disposal of colonial lands, for two reasons: firstly, such government intervention was necessary in order to realize colonists’ own interests and, secondly, the colonization of the present generation would unintentionally affect the interests of the world or of future generations. On these two grounds, Mill came to agree with “the Wakefield system of colonization” (Mill [1848] 1965, 958): the proposal “to check the premature occupation of land, and dispersion of the people, by putting upon all unappropriated lands a rather high price, the proceeds of which were to be expended in conveying emigrant labourers from the mother country” (ibid., 959; cf. ibid., 921).<sup>14</sup> Focusing on the first of the grounds, Mill addressed

<sup>13</sup>The first and second editions of the *Principles* said “select” instead of “demand.”

<sup>14</sup>As for the case of colonization, “the unparalleled amount of spontaneous emigration from Ireland” was reflected in the third and subsequent editions of the *Principles* (Mill [1848] 1965, 967; cf. ibid., 194–195, 376–379, 865–866). On this point, see Hollander (1985, 753–758).

the “[c]ases in which public intervention may be necessary to give effect to the wishes of the persons interested” (ibid., 956). While these cases can be regarded as the prisoner’s dilemma analyzed in modern economic theories, it appeared to Mill that “the attention of political economists [had] not yet been sufficiently drawn” to them (ibid., 956). According to the *Principles*, “[t]here are matters in which the interference of law is required, not to overrule the judgment of individuals respecting their own interest, but to give effect to that judgment: they being unable to give effect to it except by concert, which concert again cannot be effectual unless it receives validity and sanction from the law” (ibid., 956). The following serves as an example: “However beneficial it might be to the colony in the aggregate, and to each individual composing it, that no one should occupy more land than he can properly cultivate, nor become a proprietor until there are other labourers ready to take his place in working for hire; it can never be the interest of an individual to exercise this forbearance, unless he is assured that others will do so too” (ibid., 959). We may note, in passing, that this point was also referred to in Mill’s *Logic* (Mill [1843] 1974, 807).

It must be noted that Mill’s assertions investigated in this section (Sect. 4) would be a part not of his political economy but of his art of government. To elaborate, although “the doctrine of Free Trade” and “the doctrine of Freedom of Trade between nation and nation” belonged to the science of political economy, Mill’s insistence upon free (foreign) trade and government intervention in education and colonization was classified as one of the precepts of the art of government which, as has been demonstrated in Sect. 3 above, was founded on his speculative politics, including political economy. Firstly, in the *Essays*, Mill showed some features of art as distinguished from science and maintained that rules and exceptions could properly be discussed not in science but in art:

However heterogeneous the causes, it [i.e. art] carries the effects of them all into one single reckoning, and according as the sum-total is *plus* or *minus*, [...] Art says, Do this, or Abstain from doing it. [...] We may [...], in talking of art, unobjectionably speak of the *rule* and the *exception*; meaning by the rule, the cases in which there exists a preponderance, however slight, of inducements for acting in a particular way; and by the exception, the cases in which the preponderance is on the contrary side (Mill [1844] 1967, 338–339).

Secondly, in the *Liberty*, Mill remarked that the subject of the eleventh chapter of the fifth book of the *Principles*, namely the proper limits of governmental action, was addressed in the art of government: “To determine the point at which evils, so formidable to human freedom and advancement, begin, or rather at which they begin to predominate over the benefits attending the collective application of the force of society, under its recognised chiefs, for the removal of the obstacles which stand in the way of its well-being [...] is one of the most difficult and complicated questions in the art of government” (Mill [1859] 1977, 308–309).

It should also be added that in the last two decades of the nineteenth century, James Laurence Laughlin (1850–1933), who was an assistant professor of political economy at Harvard University (and became the first Head Professor of it at the University of Chicago in 1892), published, as “a text-book for colleges,” an abridged and annotated

edition of Mill's *Principles*,<sup>15</sup> and in this edition, the eleventh chapter of the fifth book of the *Principles* was entirely omitted. On this point, Laughlin argued: "The chapters on Land Tenures, the English currency discussion, and much of Book V, on the Influence of Government, have been simply omitted" (Mill and Laughlin [1884] 1893, iv). Barber (2010, 287) points out that Laughlin emphasized "Mill's comments on the virtues of free markets" and systematically purged "his sympathetic references to the state as an agent for social amelioration." Added to this, it is likely that Laughlin endeavored to focus on those parts of the *Principles* which mainly related to political economy, because he believed that "the omission of much that should properly be classed under the head of Sociology, or Social Philosophy,<sup>16</sup> would narrow the field to Political Economy alone, and aid, perhaps, in clearer ideas" (Mill and Laughlin [1884] 1893, iii–iv).

## 5 The Long-Term Aspect of the *Principles*: The Possibility of Human Improvement

According to the *Principles*, society had changed significantly since its first edition was published. Mill had been influenced by the change and, in the third and subsequent editions of the *Principles*, remarked that the association of laborers among themselves came to have a stronger possibility of being widespread among the general public. The preface of the third edition mentioned:

The chapter on the "Futurity of the Labouring Classes" [i.e. the seventh chapter of the fourth book of the *Principles*] has been enriched with the results of the experience afforded since this work was first published, by the co-operative associations in France. That important experience shows that the time is ripe for a larger and more rapid extension of association among labourers, than could have been successfully attempted before the calumniated democratic movements in Europe [...]. I have endeavoured to designate more clearly the tendency of the social transformation, of which these associations are the initial step [...] (Mill [1848] 1965, xciii).

In the third and subsequent editions of the *Principles*, Mill focused not only on the association of laborers with capitalists but also on that of them among themselves. The former was the association "which can exist between a capitalist as chief, and workpeople without a voice in the management," and the latter "the association of the labourers themselves on terms of equality, collectively owning the capital with which they carry on their operations, and working under managers elected and removable by themselves" (ibid., 775). Mill's expectation was as follows: "perhaps finally in

<sup>15</sup>According to Laughlin, Mill's *Principles* "yet remains the best systematic treatise in the English language" (Mill and Laughlin [1884] 1893, 23).

<sup>16</sup>Related to the word "Sociology," Mill remarked in the *Logic*: "The Social Science, [...] by a convenient barbarism, has been termed Sociology" (Mill [1843] 1974, 895).



all” cases,<sup>17</sup> “the relation of masters and workpeople will be gradually superseded by [...] association of labourers among themselves” (ibid., 769).

Mill pointed out, in those editions of the *Principles*, that the association of laborers could make people public-spirited: “[...] if public spirit, generous sentiments, or true justice and equality are desired, association, not isolation, of interests, is the school in which these excellences are nurtured” (Mill [1848] 1965, 768). In particular, the association of laborers among themselves would not only bring material benefits such as an increase in the productiveness of labor<sup>18</sup> but also cause “the moral revolution in society” (ibid., 792). Mill described this revolution as follows: “the healing of the standing feud between capital and labour; the transformation of human life, from a conflict of classes struggling for opposite interests, to a friendly rivalry in the pursuit of a good common to all; [...] and the conversion of each human being’s daily occupation into a school of the social sympathies and the practical intelligence” (ibid., 792). Similarly, the *Autobiography* also focused on the educational effect of the association and indicated that such an association could turn selfish persons into public-spirited ones: both laborers and employers “must learn *by practice* to labour and combine for generous, or at all events for public and social purposes, and not, as hitherto, solely for narrowly interested ones” (emphasis added; Mill [1873] 1981, 239). On the education by practice, Mill also remarked: “[...] we welcomed with the greatest pleasure and interest all socialistic experiments by select individuals [...], which [...] could not but operate as a most useful education of those who took part in them, by cultivating their capacity of acting upon motives pointing directly to the general good [...]” (ibid., 241).

As has been examined in Sect. 4.1 above, Mill denied the universal selfishness of people and admitted the possibility of a different state of society from so-called capitalism. In other words, Mill “regarded all existing institutions and social arrangements as being [...] “merely provisional”” (Mill [1873] 1981, 241). In Mill’s opinion, this view, as well as his positive attitude towards “the stationary state of capital and wealth,” distinguished him from “political economists of the old school” (Mill [1848] 1965, 753–754). According to the *Autobiography*, although in the days of his most extreme Benthamism Mill had seen “little further than the old school of political economists into the possibilities of fundamental improvement in social arrangements,” Mill came to look “forward to a time when society [would] no longer be divided into the idle and the industrious [...] and when it [would] no longer either be, or be thought to be, impossible for human beings to exert themselves strenuously in procuring benefits which are not to be exclusively their own, but to be shared with the society they belong to” (Mill [1873] 1981, 239). As for this change in his opinions, Mill argued: “In the *Principles of Political Economy*, these opinions [i.e.

---

<sup>17</sup>The word “perhaps” was added in the fifth edition of the *Principles*.

<sup>18</sup>On the relationship between the association and the increase in the economical productiveness, see also the first book of the *Principles* (Mill [1848] 1965, 183–184).



Mill's new opinions] were promulgated, less clearly and fully in the first edition, rather more so in the second, and quite unequivocally in the third" (ibid., 241).<sup>19</sup>

It must be remembered, however, that Mill, even in the third and subsequent editions of the *Principles*, viewed the state of society based on people's public-spirit as unrealistic for the time being, and endeavored to improve the social state that was based on their self-interest. According to the *Autobiography*, as has been stated in Sect. 2.2 above, Mill did not "overlook the folly of premature attempts to dispense with the inducements of private interest in social affairs, while no substitute for them has been or can be provided" (Mill [1873] 1981, 241). In the third edition of the *Principles*, Mill remarked: "The only objection [against Socialism] to which any great importance will be found to be attached in the present edition, is the unprepared state of mankind in general, and of the labouring classes in particular; their extreme unfitness at present for any order of things, which would make any considerable demand on either their intellect or their virtue" (Mill [1848] 1965, xciii). In Mill's opinion, the subject of an ultimately desirable state of society was an open question: the people of the present "[were] not competent to decide" a question "[w]hether [...] individual property in some form [...] or community of ownership in the instruments of production and a regulated division of the produce, will afford the circumstances most favourable to happiness [...]" (ibid., xciii). Mill regarded "the object to be principally aimed at in the present stage of human improvement" as "not the subversion of the system of individual property, but the improvement of it, and the full participation of every member of the community in its benefits" (ibid., 214; cf. ibid., 985–987).

It may be worth referring again to Laughlin's edition of the *Principles* and pointing out that he entirely rewrote the seventh chapter of the fourth book of the *Principles* (i.e. the chapter examined in this section). To explain further, in his edition of the *Principles*, Laughlin mainly focused on not the moral or educational but the economical aspect of the association (Mill and Laughlin [1884] 1893, 518–533). According to him, "[i]t has seemed necessary to make the most additions to the original treatise under the subjects of the Wages Question; [...] of the *Future of the Laboring-Classes* (*in which the chapter was entirely rewritten*); and of Protection" (emphasis added; ibid., iv). As has been pointed out in Sect. 4.2 above, his edition limited its target to those parts of the *Principles* which related to political economy; and regarding this point, he made the following remark on Mill's *Principles*: "With a deep sympathy for the laboring-classes," Mill "was tempted into the field of sociology in this book, although he saw distinctly that political economy was but one of the sciences, a knowledge of which was necessary to a legislator in reaching a decision upon social questions" (ibid., 22–23).

---

<sup>19</sup>Related to this point, it may be worth mentioning that Mill, in the *Logic*, quoted from his article entitled "Miss Martineau's Summary of Political Economy" (1834): "[...] it has been a very common error of political economists to draw conclusions from the elements of one state of society, and apply them to other states in which many of the elements are not the same" (Mill [1843] 1974, 903–904; quoted from Mill [1834] 1967, 225–226). As for this article, Dome (2004, 173), for example, commences the chapter on Mill by quoting from it.

## 6 Concluding Remarks

Let us summarize the main points that have been made in this chapter. Firstly, Mill was an economist who explicitly threw light on the assumption of political economy. In other words, he systematized and refined political economy as a science and distinguished it from physical and other mental or social sciences. According to Mill, political economy presupposed that human beings always tried to obtain the greatest amount of wealth with as small a quantity of labor and abstinence as possible. This led Mill to recognize the limits as well as the importance of political economy, and he subsequently attached great significance not only to political economy but also to other sciences, in addition to a practical knowledge of the actual world in order to put forth propositions for the guidance of mankind.<sup>20</sup>

Secondly, the self-interested actions of individuals were not essential elements of the assumption of Mill's political economy. His political economy did not necessarily presuppose selfish people who endeavored to acquire wealth *for themselves*. In other words, although it was likely that Mill, in the *Principles*, mainly focused on such selfish persons, the assumption of his political economy could be that human beings always attempted to obtain the greatest amount of wealth *for the benefit of all concerned*. It was the desire for wealth rather than self-interest that Mill explicitly focused on when he defined political economy.

Thirdly, those people who always desired the greatest amount of wealth for their own interests and who were mainly presupposed by Mill's political economy did not reflect his ideal. In his opinion, in the latter half of the nineteenth century, the association of laborers among themselves could gradually cultivate their minds and foster their public spirit in the long run and, finally, they could and should try to obtain desirable things other than wealth for public benefit. This long-term view that individuals' preferences could and should be changed has been emphasized, for example, by Matsui (2006) and recently by Barker (2018, 53–56) and, in fact, differentiates Mill's *Principles* from present famous textbooks on modern economics (e.g. Acemoglu 2009; Mankiw [1991] 2019).

Fourthly, the aforementioned individuals who only desired wealth for their own benefit failed to precisely correspond to real human beings in Mill's lifetime. He argued, however, that a political economy which mainly presupposed those people could be one of the foundations of what he called practical politics or the art of government and, consequently, of the actual guidance of mankind. According to Mill's art of government, in the current state of European countries, governments generally ought not to interfere in ordinary economic activities, while they likely should in the cases of education and colonization. For the time being, in his lack of frustration, despair or confirmation of his ideal, Mill depended on the supposition that most people tried to obtain wealth for their own interests, and he attempted to point out the surest way of eradicating poverty and reconciling the self-interested actions of people with the greatest amount of public welfare. Mill was not only "an

---

<sup>20</sup>For further details of Mill's definition of political economy, see, for example, Persky (1995) and Hinnant (1998).

excellent professor of abstract science,” including political economy (Mill [1844] 1967, 333), but also endeavored to become “a reformer of the world” (Mill [1873] 1981, 137).

## References

- Acemoglu, D. 2009. *Introduction to modern economic growth*. Princeton: Princeton University Press.
- Barber, W.J. 2010. J. Laurence Laughlin. In *The Elgar companion to the Chicago School of Economics*, ed. R.B. Emmett, 287–290. Cheltenham and Massachusetts: Edward Elgar Publishing.
- Barker, C. 2018. *Educating liberty: Democracy and aristocracy in J. S. Mill's political thought*. Rochester: University of Rochester Press.
- Blaug, M. 1958. *Ricardian economics: A historical study*. New Heaven: Yale University Press.
- Claeys, G. 2013. *Mill and paternalism*. Cambridge: Cambridge University Press.
- De Marchi, N.B. 1974. The success of Mill's Principles. *History of Political Economy* 6(2): 119–157.
- Dome, T. 1999. Bentham and J. S. Mill on tax reform. *Utilitas* 11 (3): 320–339.
- Dome, T. 2004. *The political economy of public finance in Britain 1767–1873*. London: Routledge.
- Hinnant, C.H. 1998. The invention of homo oeconomicus: A reading of John Stuart Mill's “On the Definition of Political Economy”. *Prose Studies: History, Theory, Criticism* 21 (3): 51–68.
- Hollander, S. 1985. *The economics of John Stuart Mill*. 2 vols. Oxford: Basil Blackwell.
- Hollander, S. 1987. *Classical economics*. Oxford: Basil Blackwell.
- Mankiw, N. G. [1991] 2019. *Macroeconomics*. 10th ed. New York: Worth Publishers.
- Matsui, N. 2006. John S. Mill: Morarisuto no Keizaigaku [John Stuart Mill: The Economics of a Moralist]. In *Shinpan Keizaishisoushi: Shakaininshiki no Shoruikei [The History of Economic Thought (The New Edition): Various Types of the Interpretations of Society]*, ed. K. Ota, N. Suzuki, T. Taka and K. Yagi, 98–111. Aichi: The University of Nagoya Press.
- Mawatari, S. 1987. J. S. Mill studies. *Annals of the Society for the History of Economic Thought* 25: 2–13.
- Mawatari, S. 1989. Book review: The economics of John Stuart Mill. *History of Political Economy* 21 (2): 405–407.
- Mawatari, S. 1997a. *J. S. Mill no Keizaigaku [The Economics of J. S. Mill]*. Tokyo: Ochanomizushobo.
- Mawatari, S. 1997b. *Keizaigakushi [The History of Economic Thought]*. Tokyo: Yuhikaku.
- Mawatari, S. 2001. Further research programs on J. S. Mill. *Annals of the Society for the History of Economic Thought* 39: 42–49.
- Mill, J.S. [1834] 1967. Miss Martineau's summary of political economy. In *Collected Works of John Stuart Mill*. vol. 4, 223–228. Toronto: University of Toronto Press. [CW, IV, 223–228].
- Mill, J.S. [1843] 1974. *A system of logic, ratiocinative and Inductive*. 2 vols. In CW, VII–VIII.
- Mill, J.S. [1844] 1967. *Essays on some unsettled questions of political economy*. In CW, IV, 229–339.
- Mill, J.S. [1848] 1965. *Principles of political economy with some of their applications to social philosophy*. 2 vols. In CW, II–III.
- Mill, J.S. [1859] 1977. *On liberty*. In CW, XVIII, 213–310.
- Mill, J.S. [1861a] 1977. *Considerations on representative government*. In CW, XIX, 371–577.
- Mill, J.S. [1861b] 1969. *Utilitarianism*. In CW, X, 203–259.
- Mill, J.S. [1865] 1969. *Auguste Comte and positivism*. In CW, X, 261–368.
- Mill, J.S. [1867] 1984. *Inaugural address delivered to the University of St. Andrews*. In CW, XXI, 215–257.
- Mill, J.S. [1869] 1984. *The subjection of women*. In CW, XXI, 259–340.
- Mill, J.S. [1871] 1988. *The army bill*. In CW, XXIX, 411–415.

- Mill, J.S. [1873] 1981. *Autobiography*. In *CW*, I, 1–290.
- Mill, J.S. and J.L. Laughlin. [1884] 1893. *Principles of political economy, abridged, with critical, bibliographical, and explanatory notes, and a sketch of the history of political economy*. New York: D. Appleton and Company.
- O'Brien, D.P. 1975. *The classical economists*. Oxford: Clarendon Press.
- O'Brien, D.P. 2004. *The classical economists revisited*. Princeton and Oxford: Princeton University Press.
- Persky, J. 1995. The ethology of homo economicus. *Journal of Economic Perspectives* 9 (2): 221–231.
- Riley, J. 1994. Introduction. and explanatory notes. In *Principles of political economy and chapters on socialism*, ed. J. Riley, vii–xlvii, 437–450. London and New York: Oxford University Press.
- Riley, J. 1998. Mill's political economy: Ricardian science and liberal utilitarian art. In *The Cambridge companion to Mill*, ed. J. Skorupski, 293–337. Cambridge: Cambridge University Press.
- Schwartz, P. 1972. *The new political economy of J. S. Mill*. London: Weidenfeld and Nicolson.
- Senior, N.W. [1848] 1987. Mill, the Principles of Political Economy. In *John Stuart Mill: Critical assessments*, ed. J.C. Wood, 36–75. vol. 2. London and Dover: Croom Helm.
- Sugihara, S. 1985. Jiyuboueki, Hogoshugi, Shokumin: Jiyuboueki-Teikokushugi no Haitai [Free Trade, Protectionism and Colonization: The Appearance of the Imperialism of Free Trade]. In *Jiyuboueki to Hogoshugi: Sono Rekishitekitenbou [Free Trade and Protectionism: From a Historical Perspective]*, ed. C. Sugiyama, 85–106. Tokyo: Hosei University Press.
- Sugihara, S. 2003. *Jiyu to Shinpo: J. S. Mill Kenkyu [Liberty and Progress: A Study on J. S. Mill]*. Tokyo: Fujiwara-Shoten.
- Varouxakis, G. 2013a. Uncelebrated trouble maker: John Stuart Mill as English Radicalism's foreign politics gadfly. In *John Stuart Mill: A British Socrates*, ed. K.N. Demetriou and A. Loizides, 126–153. Basingstoke and New York: Palgrave Macmillan.
- Varouxakis, G. 2013b. *Liberty abroad: J. S. Mill on international relations*. Cambridge: Cambridge University Press.
- Winch, D. 1965. *Classical political economy and colonies*. London: London School of Economics and Political Science.

# Léon Walras on Human Nature and His Social Reform Plan



Satoshi Takahashi

**Abstract** The purpose of this chapter is to clarify the relationship between Léon Walras' notion of human nature and his reform policies toward a unique socialism. First, he divides human nature into physiologic nature and psychologic nature. Concerning the self-love included in the latter, he distinguishes egoistic self-love from that of self-preservation. Then, he deduces from human nature, four social categories: Industry, Mores, Science, and Art, and two economic agents: the individual and the State. The reform policies in the category of Mores are composed of the nationalization of land and the abolition of taxes, and those in that of Industry are composed of state interventions to provide public goods like education and anti-monopoly policies to maintain free competition. Finally, Walras expects egoistic self-love, along with the private landownership and monopoly to disappear when his reform policies are instituted.

**Keywords** Léon Walras · Human nature · Self-love for self-preservation · Nationalization of land · State intervention

## 1 Introduction

The aim of this chapter is to demonstrate that there is a relationship between Léon Walras' notion of human nature and his social reform plan. Walras is the founder of general equilibrium theory, and is one of the most eminent economists in the history of economics. His *Elements of Pure Economics (EPE)* has been viewed as describing the consequences of selfish behaviors of home economics "to maximize satisfaction of wants" (Walras 1954, p. 15). His economics has also been considered to be a mathematical expression of Adam Smith's invisible hands, and to provide a theoretical foundation of the *laissez-faire, laissez-passez* policies.

However, it is relatively unknown that Walras called himself a socialist. Unlike his traditional image, Walras was very concerned about the poverty problems of his day,

---

S. Takahashi (✉)  
Meiji University, Tokyo, Japan  
e-mail: [satotaka1106@meiji.ac.jp](mailto:satotaka1106@meiji.ac.jp)

calling them the “Social Problem.” He sought to solve the Social Problem through unique reform policies inherited from his father, August Walras. As a socialist, Walras criticized French economists who unconditionally praised *laissez-faire, laissez-passé* and; he was firmly convinced that his task was to give his father’s policies a mathematical foundation. Contrary to the economists and socialists he criticized, Walras named his reform plan “a sound society” (Walras 2010, p. 157) or “human and scientific socialism” (Walras 1987, p. 511).

Walras’ reform policies are presented in his *Studies in Social Economics (SSE)* and *Studies in Applied Economics (SAE)*. *SSE* is a volume of collected papers which consists of three parts: social philosophy, including lectures on human nature, mathematical theory for land nationalization, and policies for tax reform. *SAE* consists of theories of production, money, agriculture, the labor market, foreign trade, stock market, etc. In particular, both books include papers, written from 1896 to 1898, in which he summarized an ideal society: “Theory of Property” (Walras 2005, pp. 135–159), “The Fiscal Problem” (Walras 2010, pp. 321–349), and “Outline of an economic and social doctrine” (Walras 2005, pp. 355–394).

A careful reading reveals that these papers have a common structure; Walras starts his argument by explaining human nature, and ends by putting forward arguments supporting his social plan. This common framework leads us to expect that the notion of human nature underlies his social reform and yet the brevity of these papers prevents us from easily following his reasoning from human nature to his socialism.

Several studies have tried to explore Walras’ social thought and policies since the publication of *Œuvres économique complète de Auguste et Léon Walras* in 1987. However, there seems to be a lack of research investigating the relationship between his ideas about human nature and his economic policies. For example, while Dockès (1996) discusses Walras’ philosophy and his economic policies in detail, their relationships remains unclear and his scope is nearly limited to *SSE*. Although Potier (1998) discusses most of Walras’ economic policies in *SSE* and *SAE*, examinations of his social philosophy are lacking. By contrast, Bee (2013) deals with Walras’ ideas on human nature in detail, but his focus is restricted solely to Walras’ philosophy and methodology. Although Walker (2004) points out that Walras deduced social conditions from human nature, his arguments only suggest that there is a relationship between them (Walker 2004, p. 119).

As per our investigation, Walras considered human nature to provide a philosophical foundation for property rights, however, it cannot be said that he succeeded in explaining exactly why his policies such as the nationalization of land and the abolition of taxes rested on human nature. Here, there remains an unsettled question to consider. How is Walras’ human nature or self-love compatible with his socialistic beliefs? In other words, can we find some logical consistency in Walras’ economic works: pure economics, social economics, and applied economics? For example, Rugina (1982) maintains that these three branches are completely independent.

In the Walrasian classification, economic science does not appear as a unity since three parts mentioned above are conceived as separate units.... In other word, there is no direct line of communication in the sense of a logical sequence among the three branches (Rugina 1982, p. 229)

**Table 1** Animal and human nature

| Nature of animals   | Nature of humans  |
|---|---|
| Physiological viewpoint<br>→ Direct and immediate satisfaction of needs | Physiological viewpoint<br>→ Division of labor and Exchange of goods and services<br>= Indirect Satisfaction of Needs |
| Psychological Viewpoint<br>→ Self-love or Self-interest                 | Psychological Viewpoint<br>→ 1. Sympathy and aesthetic sense<br>2. Understanding and reason<br>3. Free will           |

Contrary to this argument, we attempt to demonstrate that the three branches are closely related to one another. Above all, we clarify a consistency between Walras’ human nature and self-love and his social policies.

In section one, we first offer a brief overview of Walras’ notion of human nature from the physiological and psychological viewpoints. In section two, we examine the two meanings of self-love in Walras: self-love for self-reservation and that which is degenerated, or related to egoism. This distinction is the key to understanding his social reform plan. In section three, we look at Walras’ classification of four social categories; Industry, Mores, Science and Art, which are deduced from human physiological and psychological nature. We then examine the functions of the individual and the State, the two major economic agents that appear in the categories of Industry and Mores. In section four, we examine various reform policies discussed in *SSE* and *SAE*. Finally, in section five, we reconstruct his unique socialism in the light of human nature and self-love.

## 2 Walras on Human Nature

In this section, we examine the notion of human nature in *SSE*.<sup>1</sup> His discussion is summarized as follows (Table 1):

Walras distinguishes between the nature of animals and humans and then divides them into two dimensions: the physiological and psychological viewpoints.

According to Walras’ explanation, a man lives a “life that purely animal”, “life that are human in the strict sense of the word” (Walras 2010, p. 68). A human’s

---

<sup>1</sup>Dockès (1996) points out that most of Léon Walras’ ideas on human nature come from his father, Auguste Walras. Léon was influenced by the natural law theory of the 17th and eighteenth centuries, such as that suggested by Pufendorf and Locke, through his father. Despite his strong objection to the social contract theory of T. Hobbes, the mathematization of the world by Hobbes, Descarte, and Spinoza considerably affected Walras’ social view. A recognition of the world through numbers that forms Walras’ method, is characteristic of the seventeenth century philosophy (Dockès 1996, p. 31).

physiological nature is an aptitude for the division of labor, whereas his psychological nature is composed of sympathetic and aesthetic senses, understanding, reason and free will.

We start by examining the human physiological nature, the aptitude for the division of labor. Although all animals, including humans, make efforts in different ways when they are hungry, thirsty, etc., and satisfy their needs, but the ways they make these efforts are different. On one hand, animals satisfy their needs immediately and directly; a hungry animal chases and devours its prey. On the other hand, a human satisfies his needs only afterwards and indirectly; a hungry human makes shoes and trades them for bread, for example (Walras 2010, p. 69); in brief, “man cannot exist without the division of labor (Walras 1996, p. 461).” These arguments imply that the market economy, that is, an economy based on the division of labor and the exchange of goods and services is an inherent, or suitable economy for the physiological human nature.<sup>2</sup>

In his view, it is the division of labor that enables humankind to be superior to other animals. According to his explanation, its superiority is due to an anatomical fact: man has an opposition of the thumb to the other four fingers of the hand, and this fact is the most conspicuous anatomical difference between man and the other animals. He hence concludes that the division of labor is a natural fact, not a matter of choice (Walras 2010, pp. 69–70). Added to this, Walras estimates that the division of labor helps human beings make infinite progress, and to become much more economically advanced.

Regarding psychological natures, Walras defines humans as persons and animals as things:

Whatever is not conscious of its self and not master of itself is a thing. Whatsoever is conscious of itself and master of itself is a person. Man, being both self-conscious and self-directing, is a person. Man alone is a person; minerals, plants, and animals are things” (Walras 1954, p. 62).

Following this distinction, Walras explains that an animal’s will is instinctive and ineluctable, whereas a human’s will is conscious and free. Unlike animals who have a blind and inescapable fate, a human, as a person, has a destination which can be pursued; he does so by knowing that he has a destination. (Walras 2010, p. 77).

From the psychological viewpoint, both animals and humans are endowed with self-love<sup>3</sup>; this is also the generic name of emotions for self-interest. The important point here is that Walras did not identify self-interest as egoism. The statements below are key to an understanding of his ideal society, which we examine in Sect. 5:

---

<sup>2</sup>“The fact of the division of labor is an inherent fact of humanity. We have never seen an isolated living man, like an animal, subsisting on his own without any relation of work or exchange with his fellows”(Walras 1996, p. 461).

<sup>3</sup>“Purely animal or instinctive inclination has self-love as its principle.... the common source of most of the emotions of the beast, welcomes pleasure as beneficial and moves away from pain as something harmful” (Walras 2010, p. 72).



I use 'self-love' here in the large and extensive sense determined by the most necessary and legitimate consideration of self-preservation, and that I do not mix this up with egoism, which is only degenerate form of it and criminal (Walras 2010, p. 72)

Moreover, Walras notes that human psychological abilities can be cultivated and steadily improved:

Superior abilities develop continuously under the influence of education and further by mutual and progressive perfection of individuals by society and of society by individuals. In this way man escapes from more and more from the animals' condition. Indeed, this pathetic little creature is man, living in an economic and moral society (Walras 2005, p. 383)

He considers that unlike animals' instinctive inclination, human nature is malleable and can be developed.<sup>4</sup> Above all, education seems to play an important role in preventing self-love from degenerating into egoism. The quotation above permits us to understand one of the features of Walras's socialism. The reason is that the plural methods of education Walras discusses, such as education, morals or customs, and cooperative associations, seem to also be applied to his future society.<sup>5</sup>

Finally, Walras also points out several human psychological abilities with which animals are not endowed. One of them is unselfish emotions, which are unknown to the beast (Walras 2010, p. 72). Unselfish emotions, such as affection for family, love of country, love of one's fellow men, sympathy for the generations that are dead and those that are still to be born, do not belong to instinct, but to the heart. These are "the ability to be moved to act with unselfishness toward to persons" (Walras 2010, p. 72). Although these unselfish emotions are the crowning feature that raises man above animal or instinctive sensibility, they are often dormant or degenerated (Walras 2010, p. 73). Added to this, understanding, reason, and free will are also a part of the psychological nature peculiar to man. In consequence, Walras calls man with these psychological abilities as *moral person* (Walras 2010, p. 79).

In addition, Walras briefly defines human nature by adopting new terms: *homo economicus*, *homo ethicus*, and *homo cœnonicus* in an article in *SAE* "Outline of an economic and social doctrine", published in 1898:

Indeed, man who has wants, divides labor and, with a view to maximum satisfaction of his wants, sells his services and buys products in such quantities that their *rareté* are reciprocal to the quantities of services and products that are virtually exchangeable, this *homo economicus* understanding, reason, and a free and conscious will, i.e. the *homo ethicus*. Both shapes is also the one who is gifted with sympathy and a feeling for aesthetic the person who is living

---

<sup>4</sup>Dockès argues that Walras supposes the permanent identity of human nature, whereas Walker disagrees with Dockès by insisting on its malleability, quoting several of Walras' original texts. See Walker 2005, pp. 115–116.

<sup>5</sup>"Man needs, to develop his faculties in order to satisfy his needs, an education, which takes a considerable period of time. Fifteen or 20 years on average are necessary to get, first an elementary education and then professional training (Walras 2010, pp. 88–89).""The child entertained by its mother opens its spirit to the feelings, the ideas, and the morals of all who lived before it, and this same child, having become a man, and when dying, hands down in his turn the fruit of his labors and of the example he has set, to the art, science, and moral principles, of those who survive him (Walras 2010, p. 99)."

**Table 2** Human nature and social categories

| Physiological viewpoint   | Psychological viewpoint  |
|---|--|
| Division of labor<br>Social category: Industry<br>Applied economics | 1. Sympathy and aesthetic sense<br>Social category: Art                                      |
|   | 2. Intelligence and reason<br>Social category: Science<br>Pure economics                     |
|   | 3. Free will<br>Social category: Mores (Property, Family and Government)<br>Social economics |

in society, the *homo cænonicus*, in short, who cultivates art, takes part in science, has moral sentiments and exercises industry (Walras 2005, p. 356)

Walras owes these new terms to *The Principles of Sociology* (1896), written by an American sociologist, F. H. Giddings.<sup>6</sup> A man, as a *homo cænonicus*, has two sides: *homo economicus* and *homo ethicus*. On one hand, a man acts as a *homo economics* who always attempts to *maximize* satisfaction of wants in the economic field; on the other hand, he acts as a *homo ethicus* in the three moral fields of art, science, and mores or customs.<sup>7</sup>

### 3 Social Categories

#### 3.1 Four Social Categories

Walras defines four social categories in his system: Industry, Art, Science, and Mores or Customs. These four categories correspond to four aptitudes of human nature, respectively: (1) the aptitude for the division of labor appears in the field of Industry; (2) sympathy and the sense of aesthetics appear in Art; (3) intelligence and reason appear in Science; and (4) free-will appear in Mores or Customs (Table 2).

The division of labor, intelligence and reason, and free will also correspond to three criteria of Walras’ social science, which are separated into three branches.

<sup>6</sup>Franklyn Henry Giddings (1855–1931) is an American sociologist who attempted to conciliate the thoughts of H. Spencer, A. Comte and G. Tarde. He taught sociology at Columbia University from 1894 to 1928. Giddings was also an economist, and became vice-president of the American Economic Association in 1885. Then, he collaborated with John Bates Clerc on the theory of distribution. Walras sent Giddings the 2nd edition of *EPE*, while Giddings sent his *The principle of sociology* to Walras in 1895. See Potier (1994, pp. 261–262).

<sup>7</sup>As Dockès notes, a man pursuing his own interests also seeks good and justice as long as they are consistent with his own interests. On the basis of human behavior, there is not only pleasure, but also a sense of duty and sympathy. Neglecting these feelings, therefore, leads to breaking apart human nature (Dockès 1996, p. 80).

Intelligence and reason are applied to establish the pure economics. Its criterion is True and its object is to describe the value in the exchange of social wealth, which appears as if it is a natural phenomenon. The division of labor is observed in the economic field whose criterion is Useful or Interest; this is the object of applied economics which makes clear the law of creation and growth of social wealth. Free will is observed in the human society and its criterion is Just; this is the object of social economics, which determines the rule of initial attribution of social wealth between the individual and the State. Walras explains that each of these three branches deals with three dimensions of human behavior: mechanical, free will, and moral (Walras 1954, p. 64).

Table 2 summarizes the correspondence between the four characteristics of human nature and four social categories. As is obvious, we cannot find any discipline corresponding to Art, whose criterion is beautiful. Hence, sympathy and the sense of aesthetics are excluded from Walras' economic system. Although a day may come when humans are guided by altruism and love for their fellows, in his view they are not there yet. As a result, when we investigate the Walrasian reform plan in *SSE* and *SAE*, no feelings other than self-love need to be taken into consideration.

### 3.2 *The Individual and the State*

First, we limit our scope to Industry and Mores, because social reform policies are implemented in this context; then, we turn to an examination of two major economic agents: the individual and the State. Both agents appear in Industry and Mores, and are also deduced from human nature. As was previously noted, Walras inherited unique thoughts of natural laws from his father Auguste. Both strongly objected to the social contract theory because they were convinced that society is not conventional and free fact, but a natural and necessary fact. Their objection results from the physiologic and psychologic standpoints of human nature. Léon Walras stated that physiologically: "the state of nature is the social state" (Walras 1996, p. 121). He also said that, from the psychological standpoint, "each moral person is an essential element of a society, and society is an essential element of any moral person" (Walras 2010, p. 99). He thus considers humans to be embedded in society by nature from both the physiological and psychological standpoints.

Walras accounts for the relationship between the individual and the State by comparing it to the relationship between a soldier and an army; there could not be an army without soldiers, nor could a soldier able to fight a conflict on his own without an army (Walras 2010, pp. 101–102). In other words, man as a moral person achieves his own destinies independently with endowed characteristics on one hand, and collectively accomplishes on the other:

His act is therefore partially imputable to him personally, but partially imputable to the community or social collectivity of which he is a member. He, thus, is partially personally responsible, but partially responsible in common with others or collectively. Now from this point of view, man's moral destiny is both individual and social. What is more, all human

destinies are in certain respects independent, but in certain other respects interlinked (Walras 2010, p. 59)

Furthermore, as an example of Walras' objection to the social contract, we mention his criticism of the absolute individualism of Emile de Girardin which Walras regarded as a corollary to social contract theory. De Girardin regarded the State as an association composed of unequal individual shareholders; hence, he asserted that the benefit and costs of public services should be allocated according to individual social positions. Walras opposed De Girardin's social view by distinguishing between a private individual in the market economy, and the public citizen in the State:

the State is a group of moral persons considered all as being equal, I want it to be compared with a community of members all having the same rights and duties.... I state that there are no rich or poor people in the State, but citizens having all the same rights and duties the general conditions. As individuals, it may be that we are living in palaces or in sheds, but as citizens under own roofs we all have the right to be under the protection of the State"(Walras 2010, p. 106).

According to Walras' social view, the authors of the social contract wrongly viewed the natural state is prior and superior to the social state, and that humans feel, think and decide only based on their sensibility, intelligence and free will. On the contrary, Walras maintained that Mores have a considerable influence on human activities and feeling; the truth is that humans feel, think, and decide not only because of individual sensibility, intelligence, and free will, but also because of the sentiments, ideas, and moral principles of society (Walras 2010, p. 98).

Finally, Walras himself saw the individual and the society as inseparable, but also considered both to be abstracted separately from each other. The individual is abstracted from the society to which he belongs, whereas general social conditions are abstracted from the people forming them. The State is a natural and indispensable agent. It is not just a sum of private individuals but represents the totality of all moral persons (Walras 2010, pp. 99–100). The State "has its own existence, which exceeds the totality of the existences of all the individuals who belong to it"(Walras 2010, p. 102).

## 4 Reform in Mores

### 4.1 Justice

We now proceed to examine the role of the individual and the State in Mores and Industry. Walras is firmly convinced that the concept of Usefulness and Justice are concordant, not contradictory. However, "in the case of incompatibility between advantageousness and justice, the latter should take precedence over the former"(Walras(2010, p. 150). Hence, we first focus on Mores whose criterion is Justice,-a society composed of property, family, and government-and then we turn to

Industry, a society with a division of labor whose criterion is Usefulness and Interest, which is called “advantageousness” here.

When Walras discussed Mores, his scope was limited to the property, that is, the distribution of social wealth. From the viewpoint of natural rights, he attempted to demonstrate the just attribution of three kinds of social wealth (personal faculties, land, and capital goods) to the individual and the State. Walras stated that “the individual and the State are two equivalent social facts and the natural rights of the State and the natural rights of the individual are equally valid” (Walras 2010, p. 102). The reason for this is that on one hand, nature had given the land to all of us, and on the other hand, it has given personal skills to each one of us (Walras 2005, p. 371). Thus the individual has the right and duty to pursue his destiny with his own skills. He owns his personal faculties, his labor service, and his wage in order to accomplish each destiny, whereas the State owns the land, its service, and the rental payments. When Walrasian reform plan is accomplished fully, the State is expected to help all people to fully achieve their own destiny with the land and its rental payments.

Walras affirmed that the natural rights to the land belongs to the State. This implies that he regarded the State as the whole of humanity:

The land does not belong to all the people of one generation; it belongs to humanity; that is to say, to all human generations... if society is a natural and necessary fact, any alienation of land is against natural law, because it wrongs future generations. In legal terms, humanity is the owner and the present generation is the usufructuary of the land (Walras 2010, p. 145)

Walras argued that only the State had the right to legitimately represent humanity and that, as its material ground, the land was suitable for the State ownership, because it was something eternal indestructible and unconsumable.<sup>8</sup> Both the individual and the State thus equally have their own natural resources necessary for their subsistence: “the State subsists on the rent of land” and “the individual remains in the full and entire possession of the fruits of his work” (Walras 2010, p. 331).

Furthermore, from the viewpoint of natural rights, taxation is a state violation of the individual’s natural rights, and conversely, private land ownership is a violation of nation’s natural rights by individuals. Thus, as a corollary of individual ownership of human faculties, labor services and their wages also belong to the individual. The same is true of the land; the ownership of the land service and its rent by the State result from the landownership of the State.<sup>9</sup>

---

<sup>8</sup>In Lausanne courses of social economics, Walras quoted a passage from a work of his father, Auguste, which supports the nationalization of land. “When it comes to property, there must be a certain ratio confessed by reason between the object and the subject of the right, there must be a certain proportion between the possessor and the possessed thing. This proportion, however, does not exist in the individual land property. The individual is not skilled at owning the land. Why that? Because the land is the permanent thing and the individual person only represents an ephemeral personality” (Walras 1996, p. 235 and Walras 1997, p. 12).

<sup>9</sup>Walras called land and personal faculties “natural capital,” while naming capital goods “artificial or produced capital”; land is not artificially produced, and persons reproduce themselves by procreation (Walras 1954, pp. 216–217). By contrast, capital goods are lacking in the foundation of natural right because they are not an intrinsic natural resource, but an artificial one. Walras hence argues that they can belong to either the individual or the State that made or purchased them.

Consequently, his thoughts on property rights are summarized in the slogan “the highest law of the organization of society”; *Liberty for the individual, authority of the State. Equality of conditions, inequality of position* (Walras 2010, p. 105). This means that the State is obliged to create initial conditions by using its own resources, whereas free competition among individuals brings them different positions in society as a consequence of exercising their own human abilities.

## 4.2 Nationalization of Land

Walras denied private ownership of land in light of natural rights. He, however, did not seem to make clear reasonable grounds for natural rights. A persuasive reason may be rooted in the thought of his father, Auguste Walras. In particular, his criticism of unearned income seems to lie in the background of Léon’s arguments. Auguste criticized unearned income based on a view of social class inspired by Saint-Simon. He made a distinction between the laboring class and the idle class; the former includes farmers, managers, merchants, liberal profession, and civil servants, while the latter, capitalists (or rentiers) and land owners. The former live by their labor, whereas the latter live by their income without any productive activity (Walras 1997, pp. 37–38).

Besides this distinction, Auguste noted that there was a significant difference in income distribution between capitalists and landlords. He demonstrates this by using his dynamic theory of a progressive society. Auguste assumed a progressive society where the population and capital goods tend to increase, and land size remains constant. As new capital goods are accumulated and multiplied in the long run, capital interest tends to fall. He hence expected that the standard of living of capitalists becomes more and more difficult to maintain, they are forced to look for a lucrative profession which guarantees them an income independent of capital interests (Walras 1997, pp. 39–40). In contrast, landlords continue to receive increasing rents thanks to an increasing population and scarcity of land (Walras 1997, p. 42). This consequence of the dynamic theory makes Auguste more generous to capitalists’ interests than to the landowners’ rent. In fact, he opposed a levy on capital interests, which tend to diminish as time passes.

Léon Walras, accepting this vision entirely, provided a brief definition of progressive society in *EPE*:

In a progressive economy, the price of labor (wage) remaining substantially unchanged, the price of land services (rent) will rise appreciably and the price of capital service (the interest charge) will fall appreciably (Walras 1954, p. 391).

He mathematized his father’s vision and applied it to the policy of land nationalization in *SSE*. The State gradually purchases land from its owners at a given price; it then leases the service of land to managers of private enterprises in the fields of agriculture, industry and commerce, while receiving from them rents consistently increasing in value as a result of economic progress. Eventually the time will come when rents

cover the total interests on public obligations, which are issued to purchase land (Walras 2010, p. 243). The State then uses this increasing rent to satisfy an increasing demand for public services such as education, national defense, and so on. Walras expected that, as time passed, this increasing rent would enable the State to make public services free of charge, and to abolish all taxes.

When the process of nationalization of land and abolition of taxes was completed, the situation would become as follows. On one hand, personal capital, its service (i.e. labor), and its income (i.e. wages) belong entirely to the individual, and on the other hand, landed capital, its service, and its income (i.e. rent) belong entirely to the State. As a result, individuals will be able to survive completely on their own wages or interest, while the State will be able to survive only on rent.

## 5 Reform in Industry

### 5.1 Moral Monopoly

Walras' reform policies in the category of Industry can be classified into two types: moral monopoly and economic monopoly. First, we examine the former. In the discussion of moral monopoly, he drew a distinction between the products of private interest, and those of public interest, according to the theory of exchange in *EPE*. The former are those whose needs are felt and measured by the individual, while the latter are those whose needs are felt and measured by the entire of community, in other words by the State. This is the case of what are now called public goods, in which the same amount of services and products are available to all citizens. National defense, justice and public education etc. are examples.

Pure economics says that, in the case of private products, it is possible for an individual to establish a decreasing demand curve for each goods. By contrast, in the case of those of public interest, this is impossible for an individual (Walras 1996, p. 485). Hence, this is an exceptional case of free competition; State intervention is required to produce the products of public interest (Walras 2005, p. 164 and Walras 2010, pp. 328–329). Moreover, the State is charged with the establishment of equal initial conditions for all citizens. As Walras notes:

These services and products satisfy needs that are the same and equal to all people; the State as a consumer calculates the amount services or products of exterior and interior security, justice, education, communication (Walras 2005, p. 164).

Obviously, Walras did not identify public interest with the sum of private interests. The public utility that Walras assumes is not the same as a collection of individual utilities, but exists as an indivisible utility exceeding the sum total of individual

utilities.<sup>10</sup> Unlike French liberal economists such as Frederic Bastiat and Paul Leroy-Beaulieu, who praised *laissez-faire* and considered self-love to be consistent with the common good, Walras asserted that the State is charged with the institutions and policies for realizing social utility. Above all, he seems to emphasize a state intervention in worker education:

While workers should not offer labor in too large quantities, they should offer work as useful as possible, so they should develop their general and professional education as much as they can....persuaded by his personal interests, the worker will do his utmost to benefit himself, and to have his children benefit from all educational facilities. It is incontestably up to the State to put these at his disposal (Walras 2005, p. 228).

## 5.2 *Economic Monopoly*

We turn now to economic monopoly, where the dividing line is drawn according to two conditions of equilibrium in *EPE*. The first condition is that the each product and service has only one price in the market, the price at which the quantity supplied equals the quantity demanded. The second condition is that the selling price of products is equal to their cost price of services (Walras 1954, p. 255).

Consumers' greatest satisfaction is realized only when these two conditions are fulfilled. By contrast, they are not fulfilled in the case of natural monopolies such as railways, water supply, gas, and so on. In order to receive the greatest profit, managers, setting the selling price, are apt to limit free competition and they may receive a difference between the selling price and the cost price. Walras hence claims that the State has to intervene to turn private monopoly into State monopoly, so that the products are supplied at cost price instead of a price that ensures maximum profit (Walras 2005, p. 166).

The second condition is fulfilled only through the price mechanism of the market, where managers pursue their own interest without harming the interests of others. Walras discussed monopoly from the point of view of "Utility" in *SAE*, whereas he deals with it from the viewpoint of "Justice" in *SSE*; he notes: "free competition does not favor a person at the expense of his counterpart" (Walras 2010, p. 136). Producers in the monopoly seek their profit at the expense of consumers' interests, because they can set the selling price of products higher than the cost price of services. As a result, this profit seeking behavior results in the consumers' utility being lower than maximum (Potier 1998, p. 388 and Van Daal 2005, p. xlii).

Economic monopoly thus causes a conflict between producers' self-love and consumers' self-love. Walras divides self-interest into two types.; one is the producer's self-interest in a monopoly, while the other is the consumer's self-interest in a free competition economy. Of course, he supports the latter.

---

<sup>10</sup>Dockès says that "services whose social utility exceeds the utilities felt by the individuals: there were externalities (Dockès 1996, p. 209).".



Besides the criticism of the natural monopoly above, Walras warns that a new type of monopoly emerges in the late nineteenth century as a consequence of the free competition under *laissez-faire, laissez-passé* policy; he explains it as follows:

A few entrepreneurs disposing of huge amounts of capital would first kill off the small ones. After that, they would fight to the death, survived by one or a coalition of two or three firms until a monopoly occurred anyway. Monopoly procures maximum satisfaction of needs only under the reservation of maximum benefit of the entrepreneur (Walras 2005, p. 218).

This quotation is from an article published in 1897 in *SAE*. Walras thought that selfishness in the France of his times contributed to rapid capital accumulation. His capital accumulation theory in *EPE* assumes that capital goods can only come from individual savings and there is no presence of individual savings without individual ownership of capital and its interest. Walras hence justifies self-love as “the motive of private interest, the stimulus of need as incentive” and regards man as “first and foremost an egoist, and then an altruist” (Walras 2010, p. 147–148).

Nevertheless, he does not accept all human actions should be led by self-love. The reason is that he distinguishes human action pursuing self-interest in two ways; one is the action compatible with others’ interests, while the other is not. The former is the one that exist only under a freely competitive market. As already noted, Walras separated self-interest for self-preservation from egoistic self-interest. If the selling price in a monopoly prevents consumers from maximizing their interests or utility, Walras has no hesitation in criticizing producers’ self-love at the expense of consumers’ interests. Meanwhile, consumers, most of them belonging to laboring class, are permitted to pursue the maximum satisfaction of their needs for self-preservation in a free competition economy.

## 6 Sound Society

We now reconstruct the Walrasian socialist economy by connecting the various reform policies that we have examined. His socialism is unique because it is not a central planning system like that of the former eastern European countries, or like western European welfare states, or like Oskar Lange’s market socialism. Walras summarizes it in the following way:

My system of the distribution and production of social wealth may be outlined as follows: the individual is intensely active in a very strong State and both are rich. The State is endowed with the most solid and secure part of social wealth: it has the land. It has all the means to carry out public services and exercise or regulate monopolies of private interest (Walras 2005, p. 378)

It is useful to clarify his meaning more explicitly. The State nationalizes the land by its natural right, and organizes a moral and economic monopoly. Walras requires the State to establish the competitive market, and to keep it working. “Free competition does not mean the absence of any *State intervention*, as such intervention is necessary

to set up and maintain free competition” (Walras 2005, p. 372). He goes on to explain the reason for this. “Landowners, workers and capitalists are inclined to set up to a monopoly of services. Entrepreneurs are inclined to lay hands on the monopoly of products”(Walras 2005, p. 372). Walras considers all individuals to have an inclination to establish a private monopoly, owing to their self-love. In other words, whatever class they belong to, they tend to prevent the members of other classes from pursuing their own interests. He thus opposes all behaviors that harms other classes’ interests to establish a monopoly.<sup>11</sup>

Walras seems to recognize that *laissez-faire*, *laissez-passé* are no longer words to claims free competition, but have turned into ones to help large companies to establish a monopoly. For example, he criticizes the monopolies observed in America: private landownership and the monopolies result in “the enormous fortunes of multi-millionaires formed in a few years and you will find speculation on the increase of the value of land and, the operation of businesses without competition”(Walras 2010, p. 157). Walras thus proposes an ideal socialist economy, which he calls “a sound society (*société rationnelle*)”:

In such a society, we should imagine the mass of capital that does not belong to the mass of capital that does not belong to the State as being, in the hands of workers, in small fraction, in the form of shares and bonds of various firms, of the bonds of cooperative enterprises, which adds to present well-being, assuring security for tomorrow, preparing for future retirement; all this thanks to private initiative (Walras 2010, p. 157)

This society is built mainly on a market economy which is composed of numerous small and medium-sized enterprises. Its features are as follows: (1) private land ownership and the private economic monopoly no longer exist because both systems bring unearned income to “parasite classes”(Walras 2010, p. 157); (2) capital goods are not created from a difference between the selling price and the cost price, as in the case of a monopoly, but from individual savings formed by an excess of wage over consumption (Walras 2010, p. 157); (3) abolition of taxes on wages, along with nationalization of land permits workers to save so they can afford to establish cooperative societies or mutual aid cooperatives.

---

<sup>11</sup>In fact, on one hand, he is in favor of the working class as long as they are consumers in the competitive, but on the other hand, his support is weakened when they form workers unions in the labor market because he regards them as a labor supply monopoly. Concerning unions resisting wage cuts, he states that even if workers’ interests appear doubtful, their right seem apparent (Walras 1990, p. 207). As an alternative to the negotiation between capital and labor, Walras recommends the creation of an arbitrage system as a provisional measure (Walras 1996, p. 582); then, he, maintaining a disagreement with strikes, expects the price mechanism in an organized competitive labor market to eventually determine the wage level (Walras 2005, p. 225).

Walras expected these associations to gradually raise workers' position to that of small capitalists.<sup>12</sup> In addition, he hopes that cooperative associations will play a moral role, that is, to become a school for the working class to learn democracy.

they (the cooperative associations) fulfill an important economic role, which is not the suppression of capital, but making a capitalist of everybody. Furthermore, they have the equally important moral role of introducing democracy into the mechanism of production and opening access for everybody to business, the true school of political activities (Walras 2005, p. 228).

From our investigation, Walras seems to assume a homo economicus who acquires enlightened self-love: "our reasoning obviously presupposes an industrious population, with an elite" (Walras 2005, p. 228). Moreover, he describes an ideal type of homo economicus as follows: "people are able to know their interests and how to pursue them, that is to say, that they must be free, reasonable persons" (Walras 2005, p. 373). Although people pursue their own interests guided by their self-love, their self-love is neither like instinctive animals, nor egoistic. As moral persons, people are not only endowed with self-love, but also with understanding, reason and sympathy. Mores, school education, and cooperative societies also have a crucial influence on human nature. As a result, people are expected to be able to guide selfish emotions in a reasonable direction.

Walras states: "free competition is, only within certain limits, a self-driven and self-regulating mechanism to produce social wealth" (Walras 2005, p. 373). In an age where *laisse-faire*, *laisse-passé* had become words used to defend monopolies and private landownership, Walras viewed competitive market as a self-regulating system, subject to state intervention and organizations like cooperative society. In other words, self-love can be let free if it is enlightened through various methods of education.

## 7 Conclusion

We have argued that Walras' socialism based on his notion of human nature. He expressed his plan by the brief phrase: "A hypothetical system of organized free competition (which is something else than simple *laisse-faire*)" (Walras 1987, p. 507).

This system can be summarized in the following way: (1) nationalization of land and abolition of taxes; their purpose is to abolish individual landownership and taxes imposed on individuals. In Walras' view, both the present systems are contradictory to the ideal property rights system deduced from his notion of human nature; (2)

---

<sup>12</sup>It may be worth mentioning in passing that in Walras' lectures on cooperative societies in 1865, he expresses an intention of turning parasite classes into workers, while raising the labor class to small capital owners. "I am charmed, for my part, to see so many people who formerly had been nothing but proprietors, to be presently in the class of the workers; and there is only one thing that I would applaud even more: it would be to see that so many people who at present are only his workers, would be able to rank later in the class of proprietors" (Walras 1990, pp. 23–24).

free public services (in particular, education) implemented in the moral monopoly and cooperative societies help workers to raise their conditions of living and to make their self-interest reasonable; and (3) State intervention in the economic monopoly leads to the elimination of private interests in railways, gas, water, communication, etc., which harm consumers' interests.

Additionally, when there is a need to judge whether the state intervention is reasonable, his pure economics is applied in the following way<sup>13</sup>: (1) nationalization of land relies on his dynamic income distribution theory; (2) the moral monopoly is judged by the measurement of individual utility; and (3) two conditions of equilibrium indicate the dividing line drawn between free competition and State intervention. Hence, we find an interdependence among the three branches of Walras' economic system.

As a consequence, Walras' three policies above realize a separation between self-love for self-preservation, and self-love for exploitation. His intention is to gradually preserve the former and abolish the latter as the society progresses. The way toward his socialist economy is expected to be completed by replacing an exploitative market economy with a sound market economy.<sup>14</sup>

## References

- Bee, M. 2013. Homo coenomicus: le caractère naturel de la division du travail et la nécessité du progrès industriel dans la théorie de l'histoire de Léon Walras. Diemer, A. et Potier, J.-P. *Léon Walras Un siècle après (1910–2010)*, ed. P.I.E. Peter Lang, 87–106.
- Dockès, P. 1996. *La société n'est pas un pique-nique Léon Walras et l'économie sociale*, Economica.
- Potier, J.-P. 1994. Classification des sciences et divisions de l' "économie politique et sociale" dans l'œuvre de Léon Walras: une tentative de reconstruction. *Économies et sociétés* 20–21: 223–277.
- Potier, J.-P. 1998. Léon Walras and applied science: The significance of the free competition principle. In *Studies in the history of French political economy*, ed. G. Faccarello, Routledge.
- Rugina, A.N. 1982. Léon Walras: The pure scientist versus the social reformer. *International Journal of Social Economics* 9(3). Reprinted in, Cunningham Wood, J. (ed.), *Léon Walras critical assessments*, 218–262. London and New York, Routledge.
- Van Daal, J. 2005. Introduction. In *Studies in applied economics*, trans. and ed. by J.V. Daal and D.A. Walker, xv–lxvii. Routledge.
- Walker, D. 2004. Les idées de Léon Walras sur la nature humaine, R. Baranzini, A. Diemer et C. Mouchot, (sous la direction de), *Études Walrasiennes*, 105–133 L'Harmattan.
- Walker, D. 2005. *Walrasian economics*. Cambridge University Press.
- Walras, L. 1954. *Elements of pure economics*, trans. and ed. William Jaffe, George Allen and Unwin.
- Walras, L. 1987. *Mélanges d'économie politique et sociale*, Auguste et Léon Walras Œuvres économiques complètes, Economica, t. VII.

<sup>13</sup>“just as we study the sun and we take care to ensure its effects when they are unpleasant, we do not see why we would not study the competition (Walras 1987, p. 303).” “The competition ought to be promoted or limited according as it is advantageous or harmful” (Walras 1987, p. 303).

<sup>14</sup>Concerning the capital formation in the sound society, Walras states as follows: “They (capital goods) are not the result of exploitation by entrepreneurs of the owners of productive factors or of the buyers of product, because the likelihood of profits and the risk of losses are correlative” (Walras 2010, p. 157).

- Walras, L. 1990. *Les associations populaires coopératives*, Auguste et Léon Walras Œuvres économiques complètes. *Economica*, t. VI.
- Walras, L. 1996. *Cours d' économie sociale et d' économie politique appliquée*, Auguste et Léon Walras Œuvres économiques complètes. *Economica*, t. XII.
- Walras, A. 1997. *La verite sociale*, Auguste et Léon Walras Œuvres économiques complètes. *Economica*, t. II.
- Walras, L. 2005. *Studies in applied economics*, trans. and ed. Daal, J.V. and D.A. Walker. Routledge.
- Walras, L. 2010. *Studies in social economics*, trans. and ed. Daal, J. v. and D.A. Walker. Routledge.

# P. H. Wicksteed on Self-interest: Resource Allocation and Social Inclusion



Yoshio Inoue

**Abstract** “Self-interest” has a variety of meanings in economics. Sometimes it literally means a selfish behavior. Sometimes it simply means an autonomous behavior. This chapter focuses on P.H. Wicksteed whose work reminds us of such a variety of meanings of self-interest, to discuss that his resource allocation theory has a social inclusive philosophy. This chapter explores the methodology of Wicksteed in his research on Dante. Through the comparison between Dante and St. Thomas Aquinas, Wicksteed showed that Dante never saw the sacred world completely separated from the secular world, nor the sacred norm coming down from up above. Instead, Wicksteed considered Dante to have thought God’s will lies hidden in secular behavior, and the social norm should be grasped out of those ordinally behavior. Wicksteed applied the same methodology to his interpretation of economic theory. He considered that the real significance of the principle of resource allocation is not in realizing an efficient allocation of resources, but in making people remind the fact that we share the same autonomous behavior with all the other people including the rich and the poor. This fact could allow poor people to recover their self-confidence on one hand and invite a variety of social inclusive policies on the other. Wicksteed’s economic philosophy reminds us that the original and real significance of market economy lies in the inclusion of all of the people.

**Keywords** Wicksteed · Dante · St. Thomas Aquinas · Resource allocation · Social inclusion

## 1 Economics and Self-interest

The aim of this chapter is to explore the possibility that the principle of resource allocation includes ideas that may lead to social inclusion.

Economics is often said to be a science of self-interest. It is also often said that economics applauds selfishness and sometimes encourages selfish behavior.

---

Y. Inoue (✉)  
Chuo University, Tokyo, Japan  
e-mail: [inoueys@tamacc.chou-u.ac.jp](mailto:inoueys@tamacc.chou-u.ac.jp)

Although modern economics, especially since the 1980s, undeniably has such a character, it is misleading to define economics in only that way. Economics has understood “self-interest” in various ways, some of which are fairly contrasting to those of today. Therefore, restoring the diversity of self-interest will add new possibilities to today’s economics.

Reflecting on the relationship between economics and self-interest, we can see that the scope of economics has not been limited to the study of self-interested economic behavior. It has also embraced “altruistic” behavior based on concern for the well-being of others. Books by Adam Smith and J. S. Mill famously show their deep interest in altruistically motivated economic behavior. Alfred Marshall also says at the outset of his *Principles of Economics*,

Attempts have indeed been made to construct an abstract science with regard to the actions of an “economic man,” who is under no ethical influences and who pursues pecuniary gain warily and energetically, but mechanically and selfishly. But they have not been successful, nor even thoroughly carried out. (Marshall 1890, p. vi)

Additionally, behavioral economics has produced many research papers on altruistic economic behavior.

Moreover, economics has many definitions of self-interest. A variety of vocabulary is used to express the self-interested mind. When we describe someone’s behavior as “selfish”, it means that the person is placing his or her own benefit over that of others. However, a person is also described as self-interested when the person decides his or her behavior based on his or her own desires, with disregard for others. Their attitude is not necessarily “selfish” because they are *not conscious of* others, and thus it is not that they are giving any priority to their interest *in comparison with* others. Self-interest, in this case, means deciding one’s behavior through one’s own decision-making process (i.e., regardless of the will or behavior of other people). Therefore, this self-interest should be understood as something close to *autonomy*, rather than selfishness.

Autonomy is an ideal image of a man that modern social philosophy admires. This concept declares that “you are the master of yourself,” being free from negative ethical aspects such as selfishness or egotism. Autonomy is a positive concept that symbolizes the better side of the modern individual and modern freedom. Additionally, modern social sciences like politics and jurisprudence have explored the possibility of a society constituted of such autonomic individuals.

Economics is no exception here. That infamous “homo-economics” is derived from the image of the autonomic individual. It is fundamentally what human beings ought to be, not what they really are. It is not a positive, but a normative human image. The presupposition of autonomic decision making under perfect competition is also derived from this same image of a man. Competitors in a perfect competition will not take their advantage in comparison with others. The agents in a perfect competition pursue their own goals independently as if there were no other rivals around. No matter how unrealistic it may seem to the eyes of modern people, the normative image of competition that was derived from the image of autonomic individuals can be considered in this way.

Therefore, if we are to criticize this image, it is useless to criticize it based on a positive analysis of real competition or individuals. Instead, we must criticize it based on a normative analysis. To do so, first, a brand-new standard of the individual to replace the autonomous one must be created. To explore the range of concepts based on different philosophies and different ideal functions coexisting causally in the concept of “self-interest” noted above, we need to clarify various functions of self-interest in economics by penetrating far into the very concept and identifying the subtle differences it contains.

This chapter focuses on the possibility of self-interest that does not necessarily contain the exclusion of others, as mentioned above. It seems almost unavoidable that self-interest as selfishness or egotism will lead to a conflict with others and the exclusion of the inferior by the superior. For many, the typical scene of such exclusion is more of the reality of the market economy. However, this is not the one and only way that self-interest on economics ought to function. If self-interest as autonomy does not always exclude others, there should be another way of thinking about the role of self-interest.

It is commonly recognized that resource allocation promotes the exclusion of inferior resources in the name of improvement of economic efficiency. If the human labor is included in resources, the exclusion of inferior resource includes the exclusion of inferior labor. Therefore, it may sound paradoxical that the principle of resource allocation can help build a foundation for social inclusion. This chapter will explore the grounds for this insight by re-reading works of Philip Henry Wicksteed. It will be made clear that Wicksteed intended to find the origin of social inclusion, which would invalidate the differences between social classes and economic conditions, in the principle of resource allocation.

## 2 Wicksteed on Aquinas

P. H. Wicksteed has received little attention, even in the history of economic thought, except that he is known as one of the founders of the marginal product theory of distribution. It is not surprising that the investigations of his works, which range from economics, theology, and the ancient Greek philosophy, to the history of medieval thought from St. Thomas Aquinas to Dante, has been scarcely attempted.

Wicksteed’s principal occupation was that of a Unitarian minister, and he never attempted to obtain a job at a university as an economist.<sup>1</sup> He was one of the leading researchers of Dante in his days, and the *Divine Comedy*, which he translated partly, especially the *Paradiso* in the Temple edition, is referred to as a standard text even today.<sup>2</sup> In addition, he had a deep knowledge of ancient Greek philosophy, including

---

<sup>1</sup>Herford (1999[1931]).

<sup>2</sup>R. W. B. Lewis uses Wicksteed’s translation of *Paradiso*, choosing from many translations. Dorothy Sayers in her translation of *Comedy* listed Wicksteed’s *Dante and Aquinas* and *From Vita Nuova to Paradiso* as basic reference books. See Lewis (2001), Sayers (1949).



Plato and Aristotle, and, as we will see later, it was his attempt to find Aristotle in the roots of Dante's work that led him to his unique understanding of economics. It will be very helpful to glance at his understanding of Dante, even if it seems far afield from economics, in thinking deeply about his understanding of economics.

Wicksteed left plenty of writings on Dante. He was engaged in the University Extension movement in those days, although he was not a university professor. His theme in the University Extension lectures was Dante for the most part.<sup>3</sup> His research on Dante was comprehensive and professional, covering Dante's literature, religion, philosophy, and biography. Wicksteed revised Boccaccio's *The Early Life of Dante* several times,<sup>4</sup> and in the prologue to the 2nd edition he criticized Boccaccio's misunderstandings about some historical facts and his biased judgement that Dante's married life was miserable.<sup>5</sup> However, our interest here is in Wicksteed's method of approaching Dante's thought and philosophy, especially his method of analyzing them.

One common practice in Dante studies is to note the influence of St. Thomas Aquinas as an origin of Dante's thought and philosophy. Wicksteed carefully investigated the relationship between Aquinas and Dante in this manner. He identified the following view as their common ground.

Aquinas and Dante are equally emphatic in their insistence on the fact of the freedom of the will. Without it the reality of the moral life disappears, rewards and punishments are impossible, and the very idea of divine justice perishes. (Wicksteed 2002[1913], p. 190)

Aquinas and Dante emphasized the existence of "free will", and tried to deeply investigate the significance of their living in medieval days, when almost all people were supposed to be subject to the disciplines of Christianity. Because of this common feature, it can be said that Aquinas established his status as a precursor to Dante.

However, Wicksteed also writes:

My own impression is that we are on much safer ground when we use the works of Aquinas at the best means of introducing us into the mental and theological atmosphere that Dante breathed, than when we assume, without special evidence, that he had actually steeped himself in the study of them and knew their exact teaching upon every point. (Wicksteed 2002[1913], p. 136)

---

<sup>3</sup>The University Extension Movement was one of the civic education movements of the late nineteenth century. Oxford University, Cambridge University, and London University led this movement and tried to spread university education to people who did not have the opportunity to receive higher education. Wicksteed was a lecturer for the movement beginning in 1887. Even after he resigned from the priesthood in 1897, he continued to be a lecturer for the remainder of his life.

<sup>4</sup>Wicksteed (1898), (1904a, b).

<sup>5</sup>In addition, Wicksteed translated the correspondence between Dante and Giovanni del Virgilio, who was a close friend of Dante's, and edited the papers of Giovanni Villani, who was a contemporary of Dante and the man who first discovered the poet. Wicksteed commented that Villani's work described Dante vividly, but sometimes lacked accuracy; for instance, Villani made a mistake on the month of Dante's death. Wicksteed also translated Karl Witte, who was one of the authorities of Dante study in the late nineteenth century. Wicksteed (1904a), pp. xiii-xiv, Lawrence and Wicksteed (1898).

While he recognized the common features between them, he was also aware of a big discrepancy. What is that discrepancy? Wicksteed described this as follows:

Aquinas, on the other hand, analyses the freedom of the will till he has analysed it away, and leaves us with the sense not that we are really and ultimately responsible for our own choice, but that we choose, even when we choose wrong, in obedience to the inevitable and unfathomable will of God. (Wicksteed 2002[1913], p. 192)

Thus after all, though Aquinas will never formulate it so, it amounts to this, that, although we do what we like, and therefore it is he, not we, on whom the ultimate choice and its responsibility rest. (*Ibid.*, p. 194)

Aquinas certainly admitted the existence of free will in the sense that people have a real feeling of deciding their own behavior at their own discretion. At the same time, though, he said it was, in fact, the invisible will of God that had created that human will, completely unnoticed by humans. Aquinas believed this was true even when people were doing evil that seemed against God's will. Thus, there was no such genuine free will existing beyond God's will. That was the conclusion of Aquinas who believed himself to be a loyal servant of God.

Why did Aquinas take such an apparently contradictory attitude? We must not fail to notice that a philosophical movement existed at that time, regarding the regeneration of metaphysics that had descended from Neo-Platonism. It was believed in the tradition of Platonism that the substance of things, which is called *idea*, exists before real things emerge. Accordingly, it was believed that the right way to directly discuss the substance of things was by asking "what is...?"

However, asking for the substance in this way was a potentially dangerous threat to the authority of the Christian Church and God. If a Christian is asked "what is God?" and answers "God is A," this statement makes it possible to assume that "God is not B." Christians must then admit the existence of B that is not God, or the existence of the field B that does not belong to God. This would induce people to imagine that there is another world outside of God's world and to believe that there will be a place that is beyond God's power. This would bring about a serious situation for the Christian Church and Aquinas.

The expression "it is A" determines the limit of things. It would logically induce the expression that "it is not B" because "it is A." As a result, the field "is not God" would appear spontaneously. Therefore, Aquinas insisted that God must not be expressed positively like "God is A," but should be expressed negatively like "God is not A, nor B." The latter kind of expression would at least not restrict the limit of God, and would therefore not have to answer clearly questions like whether there is a domain that does not belong to God. The following passage comes from these examinations.

Now God does not belong to any kind or class at all. And therefore it is impossible to "distinguish" God from that which is not God by positive definition. We can only relate him to other things negatively by saying that he is not anything that they are. (Wicksteed 2002[1913], p. 105)

However, people who ask “that question” would not be persuaded by such an answer. Questions like “what is A?” or “what is God?” did not disappear in his time. That is why Aquinas had to admit that there was free will to ask “what is A?”

The questions descended from the freedom of will posed a kind of dilemma for Aquinas and the Church because the more they answered them, the more doubt their people had regarding their faith. For Aquinas, the field of God must be beyond the reach of human reason. It was not acceptable for him that a man understood the world of God while living in this world. After all, it was up to the power of the Church whether religious authorities could maintain such negative answers to questions that arose from people’s freedom of will. Aquinas was a man who strived to the end to defend the authority of God and the Church.

Such was the image that Wicksteed grasped of Aquinas, which can be read in the passages like this:

But he [Aquinas] nowhere, as far as I am aware, says that it is the Christian philosopher’s task to deduce these truths from Scripture. He must indeed prove that they are Scriptural, but he seems to have them in his possession already, and to know both that they are true and they are Scriptural. (Wicksteed 2002[1913], p. 107)

And in one of his less known works at any rate Aquinas definitely asserts that no man’s interpretation of Scripture, however well he may be qualified, can be taken against the judgement of the universal Church. The truths of faith are proved by Scripture, but the Church declares what the truths that Scripture proves are! (*Ibid.*, p. 109.)

Such an attitude was not acceptable to Dante.

### 3 Wicksteed on Dante

The basic stand point of Dante was very different from that of Aquinas. Wicksteed saw the basic stance of Dante as follows.

These instances are enough to show that while Dante habitually moved within the circle of scholastic ideas, he did not allow it to confine him when his own thought or his poetic vision broke away from its limitations. (Wicksteed 2002[1913], p. 147)<sup>6</sup>

Aquinas’s way of thinking, or his Scholastic usage of words, had become almost the standard of intelligence in Dante’s days. Dante’s usage of words inevitably reminds us of Aquinas. This does not, however, necessarily mean that Dante’s thinking was completely descended from Aquinas, or strictly restricted to the Scholastic philosophy.<sup>7</sup> It should be reminded that the *Divine Comedy* was not written in the Latin language, but in the Florence dialect that Dante spoke in his daily life.

<sup>6</sup>See f.n. no.7 for “instances”

<sup>7</sup>It will indeed be difficult to prove how far Dante placed himself from Aquinas. However, Wicksteed noted as a side proof that the order of angels in Dante’s *Convivio* was different from that in Aquinas’ *Summa Contra Gentiles*. The difference in the orders was also confirmed in *Comedy*, and Wicksteed said this was impossible if Dante was loyal to Aquinas. Wicksteed, *op.cit.*, p. 135.

What was the characteristic thinking of Dante? Wicksteed made a clear statement on this.

Every student of the “Comedy” has noticed the systematic parallelism between the sacred and secular examples of virtue and of vice, and Dante’s steady assertion of the intrinsic worth of the secular life. (Wicksteed 2002[1913], p. 134)

For Aquinas, the sacred world must be completely separated from the secular world. The secular reason that is inclined to ask, “what is A?” can never reach the sacred world. The sacred transcends secular reason, and the human being in the secular world can only faintly guess at a sign that the sacred sometimes seems to intimate. The sacred and the secular must be essentially different worlds from each other, according to Aquinas.

Dante did not doubt the sacred world’s existence, but he did not accept that the sacred world is a world that human beings in the secular world can never touch or feel. If the sacred world could never be touched or felt by anyone, people in this world (i.e., people who are living), could never have a premonition of the gospel nor tremble with the fear of hellfire. When Dante was thinking of Beatrice, however, he was certainly conscious of the joy of supreme bliss, and in the political struggle in which he himself was caught up, he certainly saw the horror of hell in the dark shining eyes of political schemers and in the sight of countless birds swarming the corpses of casualties of war.

The sacred and the secular were not so completely separated for Dante. It was not necessarily true for Dante that people could not even have a glimpse of heaven filled with bliss while they were living in the secular world (i.e., while they were alive). In the same way, it was not true that people never had to worry about being burned by hell fire so long as they were living in this world. In fact, while Dante was wandering through hell and purgatory led by Vergil, he was still alive in the secular world! He reached heaven after wandering through hell and purgatory while being alive and had a moment of blessed reunion with Beatrice. In other words, he was able to touch the sacred world while he still belonged to the secular world. Aquinas would not accept such a story.

Although the basic concept of the *Divine Comedy* was like that, Dante did not describe heaven or hell as a world that would literally appear in this world. The sacred world does not *descend from* the sky, nor *ascend from* the ground. It was not the sacred world that came into the secular world, but it was Dante living in the secular world who had to go into the sacred world. It was in the secular world that the “entrance” into the sacred world was hidden. It is very important to acknowledge this order of things.

If the sacred world comes to the secular world, then the sacred world must have existed before the secular world started to exist. There would be little difference from the Platonism that presupposed the preexistence of the sacred world independently of human experiences. Dante did not share such an idea. How, then, did Dante relate the sacred world to the secular world? Wicksteed expounded on this issue below.

And this is not a secularising of the spiritual but a spiritualising of the secular order of things. (Wicksteed 2002[1913], p. 134)

Dante attempted to build a bridge between the sacred and the secular, but he did not try to make the spiritual or the sacred descend to the secular. His intention was not to change the sacred world into easily understandable concepts by comparing it to secular events. What he intended was “a spiritualizing of the secular order of things” (i.e., inducing sacred connotations from secular incidents). His purpose in this was to reconsider secular lives deeply and find the elements of ordinary life that would lead people either to heaven or hell, regardless of their secular success or failure. It was his fundamental position to regard the idea of heaven and hell as the *abstraction* of the experience of human being.<sup>8</sup>

The elements leading people to heaven or hell could not be identified by only obeying the teachings of the Church. Most of the monstrous figures Dante encountered in hell were pious believers of Christianity while they were alive. Why, then, were they in hell? What decides your destination should be sought independently of Christian dogma and mundane morality. This is the meaning of the spiritualizing of the secular. Why does one behavior lead people to bliss and another to judgment? People had to think about this themselves. That was the reason that Dante thought people needed “freedom of will.”

Dante tried to investigate whether the substance of good and evil was hidden in people’s ordinary behavior. The reason Dante emphasized the importance of the “intrinsic worth of secular things” was that he supposed that sacred ideas were latent in secular things, like an embryo. It was just the opposite of some kinds of secularism, because the latter opposed sacred ideas and celebrated secular lives.

## 4 Wicksteed on Aristotle

Why did Dante think there was an embryo of a sacred idea in secular experiences? Wicksteed found some Aristotelean elements in Dante’s thought. Wicksteed’s *Dante and Aquinas* was a work that originally attempted to make it clear not only that Dante and Aquinas were under the influence of Christianity, but also that they were far more under the influence of ancient Greek philosophy, especially that of Plato and Aristotle. This is obvious in the first pages of his book.

The student of Dante who has firmly grasped the conceptions of form and matter (or material) has, for his immediate purpose, “captured the very citadel of philosophy.” Till then he “sees men as tree[s] walking.” And in clearing his vision and securing the desired position of command he will find no more potent ally than S. Thomas Aquinas. But Form and Matter

---

<sup>8</sup>The people who were being burned with hell fire in *Inferno* and *Purgatorio* were the real-life people with whom Dante had a direct acquaintance. Moreover, some of them were still alive when Dante was writing *Comedy*. This means that Dante did not want to lay imaginary punishments for their secular sin. Instead, he probably wanted to show what kind of “evil” might have been contained in the behavior they displayed while they were alive, through poetic or symbolic descriptions.

are technical terms of Aristotelian philosophy, and so for the key to Dante and Aquinas we must go to Aristotle. (Wicksteed 2002 [1913], p. 1)<sup>9</sup>

As is generally understood, Plato's *idea* exists before things emerge, and real things emerge as shadows of idea. Human beings can feel and acknowledge only up to the shadows. They cannot see nor touch ideas directly. But, as we obtain the concept of "triangle" by seeing every concrete triangle, we obtain a confidence regarding the existence of idea through the shadows of idea. That was the object of Plato's thought.

Aristotle disagreed with Plato's way of thinking that idea existed a priori. Aristotle thought that it was real things having concrete shape and form that had existed before human beings had an idea about them. This was clear because figures and things shaped in a triangular form existed before people had a confidence of idea of a triangle.

Therefore, things cannot exist without concrete materials, or "matter" in Aristotle's terminology, from which they are constituted. It is a human ability that recognizes the common nature or elements in things that discriminate between their own identity and concrete materials. As a result, people can determinate the substance of things, or "Form" in Aristotle's terminology, via abstraction from concrete things. Form emerges through the human activity of recognition. Although form can be said to be basically equivalent to idea, form does not preexist and come into things. To the contrary, things with matter exist before any ideas, and form is an abstracted idea that is found within and constructed from concrete things. Wicksteed writes about these things as follows.

But Plato is always trying to get at something *behind* the concrete and Aristotle to get at something *in* it. (Wicksteed 2002[1913], p. 19)

The Platonic "kinds" or *ideas* exist apart from individual things and are the perfect prototypes of which they are the imperfect imitations or reflections; the Aristotelian "kinds" or *forms* are abstractions of the human mind that have no actual existence except in transient and concrete individuals. (*Ibid.*, p. 19)

It will easily be seen that as Plato is looking for something absolute and Aristotle for something relative. (*Ibid.*, p. 19)

Wicksteed's understanding of Plato and Aristotle might seem too simple in light of the present standards of research. However, here we should not note his historical limitation, but instead make clear his way of thinking about economics, especially about the result of self-interest. This should be done by referring to his way of thinking about philosophy.

Wicksteed understood Aristotle as mentioned above, and regarded Dante as the successor of Aristotle. Accordingly, it should be possible for us to understand the relationship between the sacred world and the secular world in terms of Aristotle's denial of the preceding *idea*. Dante, too, did not hold the conception that the sacred

---

<sup>9</sup>Wicksteed said we should pay attention to Plato as much as Aristotle in the sentence right after this. Plato's "reminiscence" will play a very important role in Wicksteed's economics, as discussed later.

world had existed before the real world, with heaven or bliss and hell or hell fire then coming down to the secular world. This was because such a vision was basically Platonic. While Aquinas may have been sympathetic to this type of approach, Dante was not.

Dante took an Aristotelian approach to these questions. He would have had the image of the form of human emotions, like happiness or fear, emerge only after emotional experiences of the matter that had already been there. Needless to say, real and material experiences belong to the secular world. A variety of behavior bring happiness to people, while they also bring extreme horror. The form of heaven and hell are derived through the abstraction of real experiences and real emotions that people have. Such was Aristotle's manner of thinking, and Dante basically succeeded him in this manner of thinking.

The shape of real happiness will vary from person to person. The shape of real fear will also vary from person to person. Despite that, we have an ability to extract and recognize common features in those individual and specific cases. We have come to call them *happiness* or *fear* in general. Dante would have thought that the ideas abstracted from those secular experiences constituted the origin of the sacred ideas. The expression "a spiritualizing of the secular order of things" would have meant the creation of sacred ideas from secular experiences.

The idea of heaven is the abstraction of the real experiences of happiness. This is why people can superimpose their personal happiness onto the idea of heaven. The idea of hell is also the abstraction of real experiences of fear and pain. This is why people can feel real horror at just the idea of hell even though its existence is not really certain. The fact that those ideas have the power to arouse emotions indicates that their origin lies in their emotional experiences. We can, therefore, trust in those ideas as something to guide through our complicated life. They store universal memories of shortsighted human behavior and their consequences. The canonicity of ideas will not have a real impact if they do not have their roots in people's own experiences.

Wicksteed would have been well acquainted with such an Aristotelian way of thinking via Dante. He was increasing his interest in and appreciation for Aristotle more and more, and he devoted his last years to translating Aristotle's *Physics*.

He learned from Aristotle that every idea has its origin in a real concrete phenomenon, even if its appearance in the real world may differ at each time. Ideas do not exist without a relation to the real world. Ideas are not just copies of the real world, and they do not represent themselves as the average of real phenomena. Ideas are what we receive as intrinsic and common elements through abstraction and reconstruction of real phenomena. Wicksteed expressed this series of activities with the word "reflection." This is evident from the sentences below.

We must therefore assume some internal "common" sense which unites and relates the data of the special senses, and so enable[s] us to receive a sense impression of an object, with all its connected and related qualities or attributes of colour, resonance, smell. This "common sense," then, constructs the concrete sense images within us. (Wicksteed 2002 [1913], p. 156)<sup>10</sup>

---

<sup>10</sup>See also f.n. no.15.

If we are conscious of the image it is only by deliberate reflection and by self-examination of our consciousness that we become so. Our mind then by means of an immaterial image within us becomes actually conscious of a material object outside us. (*Ibid.*, p. 157)

Reflection is not an action valid only for religious concepts. It is valid for epistemic activities in general and can also be useful in understanding the order of the secular or things in everyday life (e.g., the economy). Therefore, this mode of thinking should have much to do with the economy as well, and it could be applied to investigating the economy.

Wicksteed performed this application in his economics.

## 5 Wicksteed on Economics

### 5.1 *Economics as Reflection*

Wicksteed's major work in economics is considered to have established the marginal product theory of distribution (Wicksteed 1992[1894]). However, as Lionel Robbins said, the marginal product theory of distribution was already "in the air" by that time. Wicksteed's contribution was basically the arrangement and coordination of various types of theories, even though he also provided some original insights.<sup>11</sup>

What Wicksteed truly did in economics was to develop the exchange theory of marginal utility into the resource allocation theory. As Robbins noted repeatedly in his *The Nature and Significance of Economics*, his thesis, "the optimum allocation of scarce resources," had its true origin in Wicksteed (Robbins 1932). As a university student, Wicksteed majored in theology, not economics. He happened to read Henry George's *Progress and Poverty* in 1879 and was deeply impressed by the book. He began to study W. S. Jevons' *The Theory of Economics* by himself, and soon presided over a private economic circle, from 1884 to 1888.<sup>12</sup> Consequently, Wicksteed had confidence that the theme of economics should be found in rational resource allocation.

Wicksteed defined the principle of resource allocation as follows in his *The Common Sense of Political Economy*.<sup>13</sup>

---

<sup>11</sup>Herford (1999), p. 232. The part relating to economics in this biography was actually written by L. Robbins. Additionally, the mathematical formula using Euler's theorem of the marginal product theory of distribution did not belong to Wicksteed himself. It was devised by A. W. Flux in his review of Wicksteed's *Co-ordination*. See Flux (1894).

<sup>12</sup>Herford (1999), chp.IV. This "circle" includes G. B. Shaw, H. S. Foxwell, F. Y. Edgeworth, Sidney Webb, and Graham Wallas. Alfred Marshall also happened to attend once. Marshall wrote to J. N. Keynes that he was criticized by Wicksteed in that the hypothesis of increasing returns was contradicted by the supposition of competitive market structure, and Marshall could not respond to Wicksteed sufficiently. This fact suggests that Marshall had acknowledged the so-called "Marshallian dilemma" before he published *Principles of Economics*. See Whitaker (1996), vol.1, p. 275.

<sup>13</sup>Wicksteed (1999[1910]). All the citations below are from the 1999 edition.



In current language “economy” means the administration of any kind of resources (time, thought, or money, for instance) in such a way as to secure their maximum efficiency for the purpose contemplated. It is administration with a minimum of waste. (Wicksteed 1999[1910], pp.13-14)

Wicksteed said clearly that the word “economy” should indicate the efficient administration of any kind of resource, from goods and money to time and thought. In other words, “economy” should mean the administration of resources in a way that minimizes waste. Additionally, Wicksteed said that:

“Political” Economy would, by analogy, indicate the administration, in the like manner, of the affairs and resources of a State, regulated as an extended household or community, and regulated by a central authority; and the study of Political Economy would be the study of the principles on which the resources of a community should be so regulated and administrated as to secure the communal ends without waste. (Wicksteed 1999[1910], p. 14)

But in all cases, whether she is spending money, helping the potatoes, pouring out the cream, or exercising a more general vigilance over bread and milk, she is engaged in the same problem of the administration of resources and she is guided by the same principles. (*Ibid.*, p. 20)<sup>14</sup>

Wicksteed emphasized that the principle of national budget management was entirely the same as that of common household management. There is no difference between the budget allocation of national defense or national education and the method of sharing potatoes among children, because they are both allocating resources. Wicksteed believed that every mother’s housekeeping concealed a general principle of allocation that could be applied to the allocation of the national budget. For Wicksteed, it was the role of economics to abstract such general principles as economic theories, or the form of economic behavior, from these daily unconscious experiences, or the matter of economic behavior. It can be said that it was a kind of “reflection” that characterized Wicksteed’s method of economics.

Reflection would be expected not only to raise form out of matter but also to determine the normative implications in the form. Dante’s heaven and hell were not just stories that such sacred worlds were existed, but they also provided a lesson for people in the secular world. We can find the same development regarding the idea of “economy” in the following.

The tendencies of modern thought and the condition of modern life have combined to sever the consideration of the administration of resources from the discussion of the ultimate ends it has in view; and it has therefore become usual to treat Political Economy as concerned with increasing the communal means rather than securing communal ends; and thou there has recently been some reaction against this tendency it is still dominant. (Wicksteed 1999[1910], p. 15)

Although the intention of this passage is not so clear, Wicksteed noticed that economics in his time was beginning to incline to the methodology of resource administration and to lose interest in the true object of economics. Furthermore, he

---

<sup>14</sup>“She” is a personal pronoun here, not the state.

seemed to be strongly concerned about the tendency of political economy to focus on increasing the volume of resources of nations. If political economy or economics is the study of efficient resource allocation, it should be supposed that the resources are *given*. Nevertheless, individuals were inclined more and more to consider political economy as not the study of *given* resources but as the study of *increasing* such resources. It seems that Wicksteed did not think this tendency was favorable.

He did not explain his reason clearly. In 1910's, though, when this book was published, the world was in the midst of high imperialism. The imperial powers, including Britain, were pushing forward to other worlds with their expansionism, exploiting resources by excluding local societies. The English historical school of economics principally supported such tendencies in policies. The Cambridge school did not support protectionism, but it did not raise any objection to external expansions of British economic interests.

It was not clear whether Wicksteed wished to criticize those movements. At the least, however, what he deemed the original meaning of "economy" was not outward expansion accompanied by the exclusion of others. For him, "economy" was allocating the *given* resources for various uses, not *increasing* the volume of resources. This interpretation was faithful to the original meaning of "economy" in the ancient Greek, which meant precisely the prevention of waste.

The rationality observed in resource allocation comes from sincere emotions like "I don't want to lose" (i.e., self-interest). The rationality that people show in allocating their salary, or even the rationality children show in using their small amounts of pocket money, are basically derived from self-interest. It should be acknowledged that the driving force of resource allocation is self-interest. However, it should simultaneously be acknowledged that allocation behavior that just shifts given resources or money from left to right does not in its essence necessitate the exclusion of someone for the benefit of the self. As discussed in detail later, this point of view is based on Wicksteed's own understanding of self-interest.

Wicksteed discovered many examples of economic principles like resource allocation in daily positive behavior like housekeeping, in the same manner of form contained in matter. Moreover, he took up the normative implications of such principles that are beyond our usual imagination. This was the basic method of Wicksteed's economics. This is also the reason why we think of his economics as the economics of "reflection."<sup>15</sup> The next epigram of *The Common Sense* might show that he thought of "reflection" as the object of economics:

---

<sup>15</sup>Therefore, if we understand common sense in this title to be general knowledge, as we usually do, we will fall into error. As we saw in the passage with f.n. no.10, Wicksteed referred to this word again in *Dante and Aquinas* and defined common sense as a function "which unites and relates the data of the special senses, and so enable[s] us to receive a sense impression of an object." However, it is still a sense impression, and is yet to become a meaningful object. The role of "reflection" is to assist in the construction of such a meaningful object. Common sense in this case seems to be the economic "common experiences" shared by everyone and, at the same time, the material from which implications must be taken. Reflection takes this role, which is exactly what this book attempts to do. Additionally, Wicksteed called the act of directing reflection to outer things "intention." Intention is the characteristic concept of the Austrian school from C. Menger to today. This indicates that Wicksteed was influenced by the Austrian school as much as Jevons. In this connection, R. S. Howey

We are all doing it; very few of us understand what we are doing.

## 5.2 *Resource Allocation and Social Inclusion*

It was more obvious in *The Alphabet of Economic Science* (1999[1888]) than in *The Common Sense* that Wicksteed thought of economics as an investigation that would make it clear that daily activities like housekeeping were actually the appearance of economic principles, and that they contained a variety of normative implications, which might even remind us of the hidden meanings of our everyday lives.<sup>16</sup>

This small book looks like just an introduction to economic theories, and most of its contents were devoted to an explanation of the differentials used in economics. Of course, it was a pioneering work for a book of his time. However, what we should find in it today is not its novelty in a mathematical sense.

Wicksteed explained the core of the marginal theory of economics as follows.

The scope and significance of this result will become more and more apparent as we proceed; but we can already see that desiredness at the margin of a unit of any commodity, expressed in terms of desiredness at the margin of a unit of any other commodity, is the same thing as the value-in-exchange (or exchange-value) of the first commodity expressed in terms of the second. (Wicksteed 1999[1888], p. 79)

This is an exact expression of the logic of  $MU_1/MU_2 = P_1/P_2$  ( $MU_i$  means the marginal utility of goods  $i$ , and  $P_i$  means the price of goods  $i$ ). However, because Wicksteed had not yet reached the idea of the marginal rate of substitution, this expression must be understood within the range of the cardinal theory. Accordingly, this expression can be translated into the form  $MU_1/P_1 = MU_2/P_2$ , which practically shows the law of equality of weighted marginal utility. As we know, the law of equality of weighted marginal utility is the fundamental principle of resource allocation.

Wicksteed took a very important step forward from this point. He described a real example to illustrate the law of equality of weighted marginal utility as follows.

The clever housekeeper has a delicate sense for marginal utilities, and can balance them with great nicety. She is always on the alert and free from the slavery of tradition. She follows changes of condition closely and quickly, and keeps her system of expenditure fluid, so to speak, already to rise or fall in any one of the innumerable and ever shifting, expanding and contracting channels through which it is distributed, and so always keeping or recovering the same level everywhere. She keeps her marginal utilities balanced, and never spends a penny on A when it would be more effective if spent on B; and combines the maximum of comfort and economy with the minimum of “pinching”. (Wicksteed 1999 [1888], p. 126)

Read literally, this may sound like an easy-to-follow example. It seems, however, that he had to point out the fact that an ordinary housekeeper, who has not received

---

suggested that it was Wicksteed who used the word “marginal utility”—not the “final utility” of Jevons—for the first time in English speaking countries. See Howey (1960), p. 134.

<sup>16</sup>Wicksteed (1999[1888]). All the citations below are from the 1999 edition.

any education in economics nor had any opportunity to manage funds, behaves as if she has always known the law. Economic rationality is underlying in every daily customary behavior, and at the same time, it conducts our behavior while we are unconscious of it. It is the role of economics as “reflection” that reminds us of this fact.

The significance of the passage above is not exhausted in this observation. It is important that the example was that of an ordinary housewife. Housewives can be considered to represent ordinary people.<sup>17</sup> Wicksteed wanted to say here that rationality in resource allocation was shared among almost all people, regardless of whether they were educated or not, or wealthy or not. This is probably clear in the following passages. Let me once again cite them, albeit in a somewhat lengthy form.

It is rather unfortunate for the advance of economic science that the class of persons who study it do not as a rule belong to the class in whose daily experience its elementary principles receive the sharpest and most emphatic illustrations. (Wicksteed 1999[1888], p. 127)

For example, few students of economics are obliged to realise from day to day that a night’s lodging, and a supper, possess utilities that fluctuate with extraordinary rapidity; and tramps who, towards nightfall, in the possessions of twopence each, make a rush on suppers, and sleep out, if the thermometer is at 45°, and make a rush on the beds and go supperless if it is at 30°, have paid little attention to the economic theories which their experience illustrates. (*Ibid.*, p. 127)

As a rule it seems easier to train the intellect than to cultivate the imagination, and while it is incredibly difficult to make the well-to-do householder realise that there are people to whom the problem of marginal utilities of a bed and a bowl of stew is a reality, on the contrary it is quite easy to demonstrate the general theory of value to any housekeeper who has been accustomed to keep an eye on the crusts, even though she may never have had any economic training. (*Ibid.*, p. 127)

For the great practical difficulty in the way of gaining acceptance for the true theory is the impression on the part of all but the very poor or the very careful that it is contracted by experience. (*Ibid.*, pp.127-128)

What do these words suggest? It must be remembered that Wicksteed was a Unitarian priest when he wrote them. Although it was not a religious book, the first passage cited above clearly shows whom Wicksteed really wanted to address from the pulpit.

What he seems to suggest is that students of economics, whose fathers were usually rich, did not belong to the class whose everyday experiences were true examples of economic theory. From a different angle, the people who Wicksteed desired to address from the bottom of his heart were poor people, whose everyday experiences were examples of resource allocation. Why did they need to listen to economists?

Wicksteed suggested more than once that the principle of resource allocation and intrinsic rationality were given to and shared among all the people, from rich to poor. We can recognize it when poor people are wondering on what they should spend their last two pence, a bed or a bowl of stew. They are just performing their daily routines. The thought circulating in their minds, though, are the same as those of a wealthy capitalist who is wondering whether to invest a million pounds, in plants or stocks.

---

<sup>17</sup>Interestingly, Wicksteed always used “she” as a pronoun for the third person singular.

The capitalist and the poor man are both thinking about how they can allocate their last money to maximize their utilities. Both are using their brain to decide how to divide their last money and not decrease their utility levels. Their manner of thinking is completely the same. This is why we can build a theory, or the form of economic behavior, that is true for everyone by observing various experiences, or the matter of daily life.

However, Wicksteed would not have wanted to comfort people *emotionally* just because the rich and poor were doing the same thing. What Wicksteed really wanted to say was probably the following. The economic gap between the rich and poor and the misery of extreme poverty were near the limit of endurance in his time, while the British Empire was at the peak of prosperity. Wicksteed grieved deeply at these problems and intended to commit himself to improve the situation directly or indirectly. For example, Wicksteed supported John Trevor who used to be his assistant at the Little Portland Street Chapel and organized the Labour Church Movement, both implicitly and explicitly. The Labour Church was one of the representative labor movements in those days and later became somewhat excessive.<sup>18</sup>

Some poor laborers became desperate under these situations and started lying down on the street or getting involved repeatedly in crimes like robbery or violence. These people believed they had already been abandoned by God. Most of all, they were convinced that they had no ability to control themselves, so they had no trust in their autonomy. The preachers of holly mercy would have been no help to them in such a condition.

However, was it true that they had really been abandoned by God? Was there no possibility of thinking about their situation in other ways? Wicksteed answered this question by suggesting that the rationality of resource allocation was shared among all people from the rich to the poor. According to Unitarianism, which denied the Trinity, rationality represents God's sign for his people. Unitarianism also taught that God's will spreads all over the secular world, incarnating in the form of rational thinking. Thus, the doctrine was tolerant toward rational thinking, and scientific and industrial activities. This was one of the reasons that Unitarianism was supported by middleclass intellectuals, including many scientists and writers.<sup>19</sup>

Considering this situation, what Wicksteed truly wanted to say might have been like the following. If you are wondering which would be better for you, a night's lodging or a bowl of stew, that is proof that you are favored with the divine will and you are not abandoned by God, even though you are struggling with poverty at this moment. Your wondering about which to choose is a sign that you are sharing God's will. You know better than anyone else that you have such a rationality of wondering.

Even the most desperate poor would not deny that they wondered which choice to take. If they were told that such a common wondering was the proof of the divine will's presence, it would have been more persuasive than giving them a hundred sermons. It is likely that Wicksteed intended to restore the poor's confidence in

---

<sup>18</sup>Herford (1999), Chap.III. For Wicksteed's activities at the Little Portland Street Chapel, see Perris (1900), pp.28–40.

<sup>19</sup>See Melnyk (2008), Chap.2.

themselves by showing such undeniable facts. Alternatively, perhaps he intended to make the poor rediscover themselves as autonomous beings by making them call forth *reminiscences* of their own rationality.<sup>20</sup>

The role of economics for Wicksteed was to give people opportunities to remind them of their autonomy. It was not an accident that Wicksteed had more interest in Jevons' insight-oriented economics than Marshall's policy-oriented economics. Being familiar with the "reflection" method would have helped him find a similar function in the economics of efficiency.<sup>21</sup>

It is possible that Wicksteed's interest in economics originated in his religious purposes. However, another implication of Wicksteed's economics is more general in nature. To put it succinctly, we can gain a basis for seeing everyone indiscriminately, regardless of differences in social status or economic condition, by showing that everyone is fully equal in sharing allocative rationality.

People who rely on their experimental feelings tend to see the difference in social class, or the difference between the rich and poor, as coming from the difference in human nature itself. It is generally known that eugenics after F. Galton was praised, particularly in Wicksteed's time to justify such feelings on scientific grounds. It would have been easy to deny these ideas as ethically dangerous and trivial. However, it would have also been difficult to overcome such deep prejudice with only a goodwill counterargument that lacked scientific or logical grounds. On the contrary, it was a clear fact that everyone was sharing the same principle of allocation. It was too simple to deny. This simple fact of everyone's sharing the same nature contradicted the feeling of people who believed in the difference in nature, and this contradiction would be the beginning of overcoming of this prejudice.

As seen above, a capitalist who wonders in which plants or stocks to invest his one million pounds will gather information on the rate of expected profit from plants and the rate of expected return on stocks as much as he can. To do this, he will not hesitate to introduce any means, such as highly developed mathematical methods, a large amount of labor, and the newest type of computers and the newest version of accounting software. The image of individuals working in a luxury office of a skyscraper symbolizes the modern elite class, and that class is the center of public envy.

Meanwhile, there might be some old and deserted pub at the back of that skyscraper. People coming there might be mostly jobless and shabby drunkards. They have about five pounds, and no one knows how they got them. They are troubled at the end of the day with the decision of whether they should spend that five

---

<sup>20</sup>Needless to say, "reminiscence" is a typical concept of Plato. However, if we use "reminiscence" less strictly in philosophical meaning, and understand it as what we do in grasping the implication from our experiences, it will not cause much harm to consider "reminiscence" as a very closely related concept to "reflection."

<sup>21</sup>The following Herford passage might be a side proof for this interpretation. "The doctrine of 'marginal utility' became in Wicksteed's hand, in fact, much more than an explanation of the economic phenomena of exchange. He found in it the key, not only to a healthy economic order, but to a healthy mentality and a healthy way of life." Herford, *op.cit.*, p. 200. Inoue(2014) investigates the ideological relationship between Wicksteed's ideas on resource allocation and income distribution.

pounds on two pints of beer or one pint of beer and one sandwich. They wonder seriously which will give them more utility, as they stand alone in front of the bar. If one compares them with the elite businesspeople hitting their keyboards in the glass building, one will see the extreme difference, as great as the difference between heaven and hell.

However, the law of the equality of weighted marginal utility tells us that the meaning of what the elite businesspeople do and what the poor drunkards do are exactly the same in essence. The elite businesspeople look for the optimal allocation of a million pounds between investments in equipment and stocks, and the drunkards look for the optimal allocation of five pounds between beer and sandwiches. They are urgently calculating how they should allocate their *fund* to maximize their utility. Both are now doing exactly the same thing.

The law of the equality of weighted marginal utility or the principle of resource allocation will reveal that people who look utterly different at a glance are performing exactly the same economic behavior once their adornments are removed. The elite and the drunkard are on equal footing in this respect, even if their appearances are quite different. If this is the case, we cannot easily maintain that the elite businesspeople should be protected by society, while the drunkard deserves to be excluded, merely because the former is in our own club (i.e., he or she is as decent as us), while the latter is not. This would be a reasonable conclusion if we stretch the implication of Wicksteed's passages cited above. Based on these observations, it can be said that Wicksteed's economics has a very strong "inclusive" nature.

Social inclusion usually means policies or thoughts that oppose social exclusion. It is usually said that economics based on marginal principles, or the theory of market economy have an "exclusive" nature because it emphasizes the importance of market selection. The real market economy certainly has the exclusive function of selection. It is possible that our drunkard used to be a business elite and was excluded from the labor market by market selection. Wicksteed understood this social exclusive function of the real market economy well and advocated an income redistribution policy based on progressive taxation, the popularization of the compulsory education system, and asset equalization, including the nationalization of land. Despite these facts, we must not overlook the fact that something that could form the basis for social inclusion is hidden in the principle of resource allocation.

We find here the point that connects Wicksteed's economics and his philosophical exploration including Dante studies. He did not advocate his idea of social inclusion *from outside* economics, such as through his religious faith or other kinds of ethics. He talked about this inclusive philosophy *from inside* the principle of resource allocation. He was not trying to correct the real result of resource allocation by way of another kind of judgment apart from economics that stresses the result of the market economy as inhuman. Instead, he suggested that the principle of resource allocation potentially contains an affirmative logic for social inclusion.

Dante did not bring the sacred idea *from outside* the secular world. It was a Platonic way of thinking and was also found in Aquinas. Dante was searching for the origins of sacred ideas within experiences in the secular world and managed to grasp them

from *inside* secular experiences. Dante named the ideas “heaven” and “hell” and tried to give them a normative power to lead people’s daily behavior.

Wicksteed made an attempt in economics that was very similar to what Dante did in the *Comedy*. He grasped the universal idea of resource allocation *from inside* the daily economic behavior of a mother’s housekeeping or a drunkard’s wondering. He then managed to grasp a normative idea to guide people *from inside* the positive theory. Needless to say, merely allocating resources does not always involve the realization of social inclusion. If positive behavior always realized normative thought, any normative thought would be needless. This is not the world in which we are living.

Normative ideas are constructed as implications of positive ideas. Social inclusion is no exception. Social inclusion is not what resource allocation brings directly into the real world. It is the potential idea of resource allocation that will do it, and we must extend and reconstruct some elements in resource allocation to reach that outcome. Accordingly, just repeating conventional resource allocation will never lead to social inclusion in the real world. The real experiences of the market economy do not realize social inclusion, as it were. However, if the normative idea is founded on the positive idea, it might be able to bring the real economy closer to the normative idea by changing the manner of real, positive behavior.

Therefore, social inclusion should not be thought of as opposed to the market economy. It should be thought of as the guiding philosophy of the real economy. The market economy might be said to be a *dynamis* of social inclusion. Economic and social policies that would be taken for social inclusion should not be anti-economic forces. On the contrary, they should be pro-economic inducements to realize *energeia*. In other words, there was nothing more unnatural for Wicksteed than a market economy demonstrating social exclusion as if it were its true nature. We should investigate here why the real market economy has departed from the social inclusive function, which should have been in the intrinsic nature of the market economy.

These arguments clearly remind us of the logic of Aristotle. For Wicksteed, housewives’ household management, drunkards’ wondering about beer or sandwiches, were all the matter of economy, and he abstracted the principle of resource allocation as the form of the economy. He simultaneously drew the form of social inclusion from the matter of the principle of resource allocation. This was basically the same style as that of Dante, who abstracted the ideas of heaven and hell from within secular experiences and expected them to lead the way of life. Economics and Dante were linked together now in Wicksteed’s observation.

Wicksteed found a philosophy leading to social inclusion within the principle of resource allocation. This provides the new insight of self-interest that underlies the formation of the principle. Economics for Wicksteed comprised “reflection” that could find marvelous implications in what we have done unconsciously. As Robbins said, that may be what economics left behind in its history.

Yet curiously enough, no aspect of his teaching has been more completely neglected. The reason is not far to seek. In England, at any rate, the average economist, secure in the tradition of an empirical approach to his subject-matter, is apt to be impatient of inquiries which linger on implications and modes of conception. (Herford 1999, p. 246)



## References

- Alighieri, Dante. 1899. *Paradiso*, trans. and ed. P.H. Wicksteed. Temple Classics.
- Alighieri, Dante. 1900. *Inferno*, trans. and ed. P.H. Wicksteed. Temple Classics.
- Alighieri, Dante. 1901. *Purgatorio*, trans. and ed. P.H. Wicksteed. Temple Classics.
- Alighieri, Dante. 1949. *The divine comedy*, trans. and ed. Dorothy L. Sayers, 3 vols. Penguin Books.
- Flux, A.W. 1894. Review to K. Wicksell, *Über und Rente nach der neurren nationalökonomischen Theorein* and P. H. Wicksteed, *Essay on the Co-ordination of the Laws of Distribution*. *Economic Journal* 4: 303–313.
- Herford, C.H. 1999 [1931]. *Philip Henry Wicksteed, his life and work*. Thoemmes Press/Kyokuto Shoten Ltd., 1999 (originally published by J. M. Dent & Sons Ltd, 1931).
- Hovey, R.S. 1960. *The rise of the Marginal Utility School 1870–1889*. Columbia: Columbia University Press.
- Inoue, Yoshio. 2014. P. H. Wicksteed ni okeru Shigenhaibun to Shotokubunpai no geni (in Japanese, The original meaning of Resource Allocation and Income Distribution on P.H. Wicksteed) In *Keizaigaku wo saiken suru: shinka keizaigaku to kotenha kachiron*, ed. Shiozawa Yoshinori and Aruka Yuji (*Rebuilding Economics: evolutionary economics and classical value theory*), Chuo Daigaku syuppan-bu.
- Lawrence, C.M., and P.H. Wicksteed (eds.). 1898. *Essays on Dante by Dr. Dukworth*. London: Karl Witte.
- Lewis, R.W.B. 2001. *Dante*. Penguin Putnam Inc.
- Marshall, Alfred. 1890. *Principles of economics*. Macmillan and Co.
- Melnyk, J. 2008. *Victorian religion: Faith and life in Britain*. Praeger.
- Perris, H.S. 1900. *A sketch of the history of the little Portland Street Chapel*. McCorquodale & Co., Ltd.
- Robbins, L.C. 1932. *An essay on the nature and significance of economic science*. Macmillan.
- Wicksteed, P.H. 1999 [1888]. *The alphabet of economic science*, Thoemmes Press/Kyokuto Shoten Ltd. (originally published by Macmillan, 1888).
- Wicksteed, P.H. 1992 [1894]. *An essay on the co-ordination of the laws of distribution*. Thoemmes Press/Kyokuto Shoten Ltd. (originally published by Macmillan, 1894).
- Wicksteed, P.H. 1898. *A provisional translation of the early lives of dante, and of his poetical correspondence with Giovanni del Virgilio*, printed for the translator.
- Wicksteed, P.H. 1904a. *The early lives of Dante, The King's Classics*. London: Alex. Morning.
- Wicksteed, P.H. 1904b. *Life of Dante*. Cambridge, MA: Riverside Press.
- Wicksteed, P.H. 1999 [1910]. *The common sense of political economy and selected papers and reviews on economic theory*, 2 vols. Thoemmes Press/Kyokuto Shoten Ltd. (originally published by Routledge & Kegan Paul Ltd, 1888).
- Wicksteed, P.H. 2002[1913]. *Dante and Aquinas: Being the substance of the Jowett Lectures of 1911*. Honolulu: University Press of the Pacific. (originally published by J. M. Dent, London, 1913).
- Wicksteed, P.H. 1922. *From Vita Nuova to Paradiso*. Longmans.
- Wicksteed, P.H., and Edmund G. Gardner. 1902. *Dante and Giovanni del Virgilio*. Westminster: Constable.
- Whitaker, J.K. (ed.). 1996. *The correspondence of Alfred Marshall, economist*. Cambridge: Cambridge University Press.

# **Self-Interest in Twentieth Century Economics**

# Otto Neurath's Theory of Felicitology and the Will to Socialization



Manabu Kuwata

**Abstract** It has been assumed that the logical empiricism of ‘the Vienna Circle’ (*Wiener Kreis*) greatly influenced on the formalization of utility and rationality in welfare economics in the 1930s. However the actual historical relationship between logical empiricism and economic science in the early twentieth century has not been still sufficiently examined in the literatures on the history of economic thought. In the 1910s, Otto Neurath (1882–1945), the leader of the Circle, was developing, as economic science, an empirical study on human welfare, titled ‘Felicitology’ that was rather different from the ‘new’ welfare economics. The purpose of this chapter is to investigate Neurath’s felicitology as ‘the other welfare economics’ and its relation to his ideas on socialization and planning. Through this investigation, I show the possibilities of the empirical and critical economics developed from the philosophy of logical empiricism in the early twentieth century, which are markedly different from the ‘rational choice’ model as a purely psychological discipline based on formalized ‘self-interests’. The analysis would also present a critical perspective on the conventional interpretation of the relationship between social sciences and logical empiricism.

**Keywords** Otto Neurath · Logical empiricism · Felicitology · *Homo economicus* · Rationality · Socialization

## 1 Introduction

In the 1930s, in the depths of the Great Depression, welfare economics experienced a major shift from old to new doctrines. Specifically, it concerned the shift of its core concepts from cardinal utility, which was tied to Benthamite utilitarianism, to the ordinal one (i.e. the ordinalist revolution). It is generally accepted that the beginning of this shift was Lionel Robbins’ *On the Nature and Significance of Economic Science* in 1932. Robbins gave economic science the explicit definition of ‘the science which

---

M. Kuwata (✉)

Faculty of Urban Management, Fukuyama City University, Fukuyama City, Japan

e-mail: [m-kuwata@fcu.ac.jp](mailto:m-kuwata@fcu.ac.jp)

© Springer Nature Singapore Pte Ltd. 2021

S. Egashira et al. (eds.), *A Genealogy of Self-Interest in Economics*,

[https://doi.org/10.1007/978-981-15-9395-6\\_9](https://doi.org/10.1007/978-981-15-9395-6_9)

studies human behaviour as a relationship between ends and scarce means which have alternative uses'. Additionally, he claimed that 'interpersonal comparison of utility' or the law of diminishing marginal utility, which characterized 'old' welfare economics, was the matter of subjective value judgements and thus excluded from economic science, because '[t]here is no means of testing the magnitude of A's satisfaction as compared with B's' (Robbins 1935, pp. 139–140, italics original). For Robbins, 'old' welfare economics, which justified the distribution of wealth from the rich to the poor based on the law of diminishing marginal utility, was only 'the accidental deposit of the historical association of English Economics with Utilitarianism' (ibid., p. 141). Therefore, he insisted that, if economics will be the better and more convincing analytical science, it should be separated from utilitarian postulates. From then on, economists sought to redefine the concept of utility to denote not pleasure, but rather a formal numerical representation of any preference ordering revealed by consistent choice behaviour. The so-called 'Ordinalism' disconnected preference satisfaction from any hedonistic feelings of pleasure or pain and away with interpersonal utility comparisons.<sup>1</sup>

Accordingly, the concept of the 'rational economic man' or *homo economicus* has also come to be separated from psychological hedonism and the motivation only provided by money gains. As a result, the *homo economicus*' rationality was understood formally as implying 'consistency in choice', in the sense that if A is preferred to B and B to C, A will be preferred to C (ibid., p. 96).<sup>2</sup> The economic agent is thus rational in the sense of aiming, under the constraint of a finite budget, to maximize preference satisfaction. It is not assumed that money and self-interest are such agent's ultimate objects. Therefore, Robbins insisted on that '[s]o far as we are concerned, our economic subjects can be pure egoists, pure altruists, pure ascetics, pure sensualists or—what is much more likely—mixed bundles of all these impulses' (ibid., p. 95).

It is notable that, after Robbins, the concept of cardinal utility and its comparison between individuals were abandoned, and welfare economics advanced towards the formalization indicated by such as Vilfredo Pareto. The question is what was the intellectual background of the formalization of utility and rationality in welfare economics in the 1930s. What has often been assumed, either explicitly or implicitly, as an important context is the logical positivism or the Vienna Circle's scientific world-conception, which greatly influenced British philosophy in the early 1930s (Walsh 1987, 1996, Chap. 7; cf. Hollis and Nell 1975). This is because it had been widely thought that logical positivism was a doctrine strictly requiring strict demarcation between empirical and normative claims, dualism of facts and values, and denying the possibility of rational argument about values or normative claims. O'Neill and Uebel pointed out that '[t]he descriptive impoverishment of welfare concepts in neo-classical economics, the rejection of the possibility of interpersonal comparisons of welfare, and the shift away from the distributional questions are taken to have

<sup>1</sup>On the more focused research on the ordinal revolution, see Hands (2010).

<sup>2</sup>This use of the term of "consistency in choice" is the way that Robbins used, but in the recent literature on choice theory, this is rather called "transitivity".

their origins in the impact of logical positivism on neo-classical economics in the 1930s' (O'Neill and Uebel 2008, p. 380).<sup>3</sup> Robert Putnam also have argued that 'the fact-value dichotomy or dualism (in a virulent form, in which ethical questions were considered to be questions of 'thy blood or mine') penetrated neo-classical economics after 1932', as the result 'impoverishment of welfare economics' ability to evaluate what it was supposed to evaluate, *economic well-being*' was happened (Putnam 2002, p. 62, italics original). According to Putnam, through Robbins 'the idea that the economist could and should be concerned with the welfare of the society in an evaluative sense was rejected, and in its place was inserted the positivist idea that such a concern was "meaningless," at least from a scientific point of view' (ibid., 54).

However, the actual historical relationship between logical positivism, or empiricism, and economics in the early twentieth century has not been still sufficiently examined in the literature on the history of political-economic thought. Specifically, the philosophy of science in the Vienna Circle was not as monolithic as generally believed, and its relationship with economics was not straightforward. Notable is the economics of Otto Neurath (1882–1945), the leader of the Circle and its only social scientist. In the 1910s, Neurath was developing, as economic science, an empirical study on human welfare, titled 'Felicitology' that was rather different from the 'new' welfare economics post Robbins. His felicitology was intended as a comprehensive theory of living conditions and life quality of humans, and closely linked to the socialist political practice surrounding his later 'socialization' in the Red Vienna (1919–1934).

The purpose of this chapter is to investigate Otto Neurath's felicitology as 'the other welfare economics' and its relation to his ideas on socialization and planning. Through this investigation, I show the possibilities of the empirical and critical economics developed from the philosophy of logical empiricism in the early twentieth century, which are markedly different from the 'rational choice' model as a purely psychological discipline based on formalized 'self-interests'. Neurath's economics was clearly not a science studying rational behaviour of *homo-economicus*, but more concrete living and happiness conditions of 'man on earth'. The analysis would also present a critical perspective on the conventional interpretation of the relationship between social sciences and logical empiricism.<sup>4</sup>

In following section, based on recent studies on Neurath's philosophy of science, I first examine his position on the relationship between science and values or politics (2). Then, focusing on his felicitology, I show that, while it criticized the concept of cardinal utility associated with Benthamite utilitarianism, it is a conception of welfare economics totally different from the 'new' welfare economics post Robbins. Neurath tried to grasp the economic agent not as *homo economicus* but as 'man on the earthly plane' or 'man as geological agent', who is embedded and embodied

<sup>3</sup>Vivian Walsh pointed out that the influence of logical positivism appears more clearly in Robbins' *On the Nature and Significance of Economic Science*, 2nd edition in 1935 (Walsh 1996, pp. 176–181).

<sup>4</sup>Neurath disliked the term 'positivism' which is associated with August Comte, and consistently used the term "empiricism" from the beginning (see Haller 1988, p. 34).

within the social and ecological relations of interdependence. From this point of view, Neurath's approach to economic science was presented as a conscious criticism of 'catallactics', or marginal utility theory, and it rather belongs to the lineage of the later substantive economics of Karl Polanyi and K. W. Kapp (3). Finally, I examine the meaning of felicitology in the context of his more famous argument on 'socialization' or 'economy *in natura*' (*Naturalwirtschaft*) in the 1920s and 1930s, which have been attacked by Austrian economists, such as Ludwig von Mises and Friedrich von Hayek (4).

## 2 Critique of Pseudorationalism

Neurath, the protagonist of the Unity of Science movement, had been regarded for a long time as an earnest advocate of 'Scientism'. His idea on 'physicalism' or 'unified science' has been identified as claim towards 'physical reductionism', and therefore thoroughly criticized by Hayek in his *The Counter-Revolution of Science* (1952). For Hayek, Neurath's consistent arguing for a planned economy through 'calculation *in natura*' (*Naturalrechnung*) in the socialist calculation debate during the post-war period also exhibited his crude 'Cartesian rationalism'. Hayek criticized as follows: 'The most persistent advocate of ... *in natura* calculation is, significantly, Dr. Otto Neurath, the protagonist of modern "physicalism" and "objectivism"' (Hayek 1952, p. 170).<sup>5</sup> However, reconsidering the Vienna Circle in the philosophy of science since the 1980s, the previous views on Neurath's thought have been radically revised:

Even a cursory glance at Neurath's work...reveals that his philosophical views were quite contrary to the popular impression of logical positivist philosophy. What people usually associate with the Vienna Circle and the Unity of Science movement is, if anything, closer to Carnap's early view: basic observational statements form a secure, common foundation; all scientific knowledge can be constructed from these by purely logical means. Throughout his career Neurath attacked as 'pseudorationalist' this reductionist picture of a single structure built upwards from the 'bricks of experience'. (Cat et al. 1996, p. 347)<sup>6</sup>

Rudolf Haller was a philosopher of science who criticized the unduly simplified Vienna Circle views and sought to correct them early on. According to Haller, in this simplified picture, the Vienna Circle (a) defended a theory of knowledge derived from the faulty assumption that all statements or all meaningful statements are, in principle, derived from immediate experience; (b) dismissed the diachrony of theories from their studies; and (c) conceived the idea of cumulative progress in science. Haller argued that all three suppositions were not only wrong, but also internally criticized from the beginning (Haller 1988, p. 35). Neurath was an outstanding immanent

---

<sup>5</sup>On the debate between Neurath and Hayek on the concepts of 'physicalism' and *in natura* calculation, see O'Neill (2006).

<sup>6</sup>Wade Hands also discussed the peculiarity of Neurath position in the Vienna Circle by focusing on the protocol statement debate (Hands 2001, Chap. 3).

critic.<sup>7</sup> Of course, to discuss Neurath's position within the philosophy of science as a whole is beyond the scope of this paper.<sup>8</sup> As such, I limit the discussion to only the important issue—the relationship between science and values or politics—for the purpose of this chapter.

The first thing to note is that Neurath was more cautious than anyone else in the Vienna Circle about continuity between facts and values. He thought we cannot avoid such situations in which value judgements play a central role in scientific theorizing itself. Okruhik mentions that, for Neurath, '[s]cience involves *conation* as well as cognition' (2009, p. 58). This idea clearly manifested in the position Neurath took regarding the famous 'protocol-sentence' debate in the early 1930s. This debate unfolded among Moritz Schlick, Rudolf Carnap, and Neurath over 'the principle of verifiability'. The key point of the protocol-sentence debate was it shed light on the fundamental epistemological differences between Schlick and Carnap, who conceived of foundations of knowing, and Neurath, who did not accept any such ultimate grounding for theories to be based on (cf. Stuchlik 2011).

Schlick and Carnap defended it is possible to obtain infallible empirical propositions constituting scientific theories based on phenomenalist language about observable empirical phenomena. A phenomenalist language refers only to one's private impression about observables. Moreover, the sentences expressed in this phenomenalist language were 'protocol sentences', an integral part of the verifiability criterion of meaningfulness. As W. V. O. Quine later criticized, there existed 'reductionism', where each one of the empirical propositions could be reduced or translated to a sentence indicating immediate experience. Conversely, in 'Protocol Statements' (1932), Neurath strongly criticized the empirical foundationalism of Schlick and Carnap, which assumed the existence of atomic experiences that 'required no verification' from which 'a world-picture or a scientific system' should be constructed. 'There is no way to establish fully secured, neat protocol statements as starting points of the sciences. There is no *tabula rasa*. We are like sailors who have to rebuild their ship on the open sea, without ever being able to dismantle it in drydock and reconstruct it from the best components' (1932, p. 92). He insisted that scientific practice was better served by observations made in a 'physicalist language' that describes publicly accessible things and events over space and time, and not private sense data. However, since 'one might always be mistaken about what objects one is seeing because of aberrant circumstances in the environment, in one's visual system, or in one's memory' (Stuchlik 2011, p.190), the sentences cast in physicalist language are always open to intersubjective corroboration and revision. In other words, such physicalist protocol sentences would not be able to secure an infallible foundation of empirical science.

---

<sup>7</sup>Interestingly, Haller pointed out that Neurath seems to have misled most people working on the history of the Vienna Circle by stressing the uniform picture of its development, as well as of its main ideas (Haller 1988, p. 13).

<sup>8</sup>Cartwright et al. (1996) provide a rather comprehensive discussion on Neurath's philosophy of science.

When a new statement is presented to us we compare it with the system at our disposal and check whether the new statement is in contradiction with the system or not. If the new statement is in contradiction with the system, we can discard this statement as unusable ('false'), ...however, one can also 'accept' the statement and change the system accordingly so that it remains consistent if this statement is added. The statement may then be called 'true'. The fate of being discarded may befall even a protocol statement. There is no '*noli me tangere*' for any statement though Carnap claims it for protocol statements. (Neurath 1932, pp. 94–95)

Neurath's view that protocol sentences are always revisable and, thus, do not provide an incorrigible empirical basis of knowing, underlaid his argument for the 'underdetermination' of theory by observational data, or what Haller calls the 'Neurath principle' (Haller 1988, p. 13; cf. Haller 1991). Haller argued that one origin of the conflicting opinions within the Vienna Circle over important issues such as including protocol statements was the existence of 'the first Vienna Circle'. This first Vienna Circle lasted from about 1907 to 1912 and was informally organized with Hans Hahn (mathematics) and Phillip Frank (physics), who Neurath met at the University of Vienna. In the first Vienna Circle, in addition to political, historical, and religious issues, there were lively discussions on the circumstances surrounding physics and philosophy (epistemology) at the end of the nineteenth century, centred on authors such as Ernst Mach, Henri Poincaré, and Pierre Duhem (*ibid.*, Chaps. 1 and 2).

Specifically, the influence of the holistic idea of knowledge from Duhem's *La théorie physique*, 1906, had on Neurath was significant. Neurath first took from Duhem's work the idea that more than one self-consistent system of hypotheses can satisfy a given set of facts and, second, that any theory testing has to do with an entire network of concepts, not with concepts that can be isolated. According to Neurath, if every proposition in a scientific theory is related to all the other propositions, we can retain its consistency either by changing the propositions which does not cohere with the system or by changing the system—this is 'Neurath principle'. When Neurath put forward his famous simile of sailors who must repair a ship while it is floating on open sea, he traced back this idea to Duhem. Neurath, who was strongly influenced by Duhem's idea of holism, came into conflict with Schilick, who inclined to the 'picture theory of meaning' of Wittgenstein's *Tractatus* (1921), and was oriented towards the foundation of knowledge through association with private direct experience.<sup>9</sup>

It is worth noting that Neurath's anti-foundationalism was already noted in his early 'The Lost Wanderers of Descartes and the Auxiliary Motive' (1913):

Each attempt to create a world-picture by starting from a *tabula rasa* and making a series of statements which are recognised as definitively true, is necessarily full of trickeries. The

---

<sup>9</sup>Haller concluded that '[b]ecause if we agree that there was a first Vienna Circle formed by a proper part of that well-known group of logical empiricists, called by Neurath *Der Wiener Kreis*, we have to acknowledge that there have been very different views on almost all important questions through all the periods of its existence: not only different views, but also views which *contradicted one another*' (Haller 1988, p. 38).

Carnap also wrote on Neurath: 'He liked to spice the talks with a lot of wit and sarcasm, criticizing the views and attitudes of others, including myself and of philosophers whom I appreciated highly, such as Schlick and Russell' (Carnap 1963, p. 24).



phenomena that we encounter are so much interconnected that they cannot be described by a one-dimensional chain of statements. The correctness of each statement is related to that of all the others. It is absolutely impossible to formulate a single statement about the world without making tacit use at the same time of countless others. Also we cannot express any statement without applying all of our preceding concept formation... We can vary the world of concepts present in us, but we cannot discard it. Each attempt to renew it from the bottom up is by its very nature a child of the concepts at hand. (Neurath 1913, p. 3)

According to Neurath, Descartes's 'fundamental error' was thinking we can make a sharp distinction between the domain of action and that of thought, and require provisional rules only for the former. By contrast, Neurath emphasized that theory is just a form of practice, and in the sphere of theory we have to depend on extra-evidential considerations in choosing what hypotheses to accept or refute. Neurath proposed calling these external motives that play indispensable roles in our choice under uncertainty, 'auxiliary motives'. In pre-modern society, instinct, superstition, belief in authority, tradition, and community feeling played the roles of auxiliary motives. For Neurath, modernity was a difficult era, characterized by the fact that it was no longer possible to rely on such things. In modern society, most people rely on their insight and are convinced that, given enough thought, one could at least determine what manner of action has the greater probability of being successful. However, this attitude is a sign of 'pseudorationalism', which cannot recognize the inevitable role of auxiliary motives in both practical action and thought.

Whoever adheres to the belief that he can accomplish everything with his insight, anticipates in a way that complete knowledge of the world that Descartes puts forward as a far-off aim of scientific development. This pseudorationalism leads partly to self-deception, partly to hypocrisy. ...The pseudorationalists do true rationalism a disservice if they pretend to have adequate insight exactly where strict rationalism excludes it on purely logical grounds. Rationalism sees its chief triumph in the clear recognition of the limits of actual insight. (Neurath 1913, pp. 7–8)

The fact that Neurath emphasized the role of ungrounded 'conations' and 'decisions' of scientists in scientific practice was parallel to his view that protocol statements are chosen in decision making and then are, in principle, modified and corrected. This was a matter (true in natural science, but more so in social science) that science and its meta-theory were closely connected with the political and social contexts. Therefore, there is a continuity that cannot be separated between 'facts' and 'values'.<sup>10</sup> In principle, everything in science is a matter of decision. Additionally, such decisions cannot all be based on logical inference, but are to be investigated historically and sociologically. Neurath recognized that social scientists, and indeed any other scientists, are 'part of the social scene' and do not enjoy 'social extra-territoriality' (1931, p. 406). This is the reason why Neurath was often considered a highly unorthodox logical empiricist thinker, anticipating the later ideas of Quine and Thomas Kuhn.

---

<sup>10</sup>For a discussion of the political and social context of Neurath and the Vienna Circles' philosophy of science, see Zolo (1989), Cartwright et al. (1996), Cat et al. (1996), Uebel (2005a), Okruhilik (2009), Stuchlik (2011), Sandner (2014). In particular, on Neurath's economic thought, see O'Neill (2003, 2006), Nemeth et al. (2007), Nemeth (2013), Poettinger (2016), Turk (2016).

Similarly, Neurath's critique of metaphysics also held clear political and social intentions from the beginning.<sup>11</sup> Regarding the 1929 manifesto of the Circle—'The scientific world conception'—, it was consciously tied to the socialist politics of Red Vienna (Uebel 2005a): 'The Vienna Circle believes that in collaborating with the Ernest Mach Society it fulfils a demand of the day: we have to fashion intellectual tools for everyday life, for the daily life of the scholar but also for the daily life of all those who in some way join in working at the conscious re-shaping of life. The vitality that shows itself in the efforts for a rational transformation of the social and economic order, permeates the movement for a scientific world-conception too' (Carnap et al. 1929, p. 305). The threat Neurath tried to overcome through a scientific world conception was related to the inadequate metaphysics that supported intolerant nationalism, fascism, and totalitarianism, and at the same time, the reality of the capitalist market order falling into a 'democracy between enemies' (Neurath 1928, p. 260), further causing a highly inequitable distribution of income, mass poverty, and irrational use of all types of resources. As discusses in the following sections, Neurath's research on the economic science of human welfare (i.e. felicitology) was developed in close association with his own auxiliary motives, that is, 'the will to socialization'.

### 3 Felicitology as the Other Welfare Economics

#### 3.1 *Anti-catalactics*

Over the approximate 10-year period from obtaining his doctoral degree (in ancient economic history) at the University of Berlin under Eduard Meyer, an ancient historian, and Gustav Schmoller, a prominent economist of the German Historical School, in 1906 until his participation in political activities in 1918 after the First World War, Neurath worked on reforming the theoretical and conceptual framework of economics based on his empiricist epistemology. Although he continued to investigate the methodological analysis of social sciences until the 1940s, his conceptual foundations of economics had already been established during the 1910s. Over that period, Neurath organized the first Vienna Circle and assimilated the ideas of the Austrian school of economics, mainly through the seminars organized by Eugen von Böhm-Bawerk. The main goal of his research during that period, as systematically shown in 'The Conceptual Structure of Economic Theory and its Foundation' (1917), was to show the limits of conventional market economics and build a broader theoretical framework (i.e. 'comparative economic theory') that 'is able to provide equal theoretical treatment to all possible forms of economic activity', including

---

<sup>11</sup> According to Carnap, '[a]ll of us in the Circle were strongly interested in social and political progress. Most of us, myself included, were socialists. But we liked to keep our philosophical work separated from our political aim... Neurath criticized strongly this neutralist attitude, which in his opinion gave aid and comfort to the enemies of social progress' (Carnap 1963, p. 23).

non-market economy (1917, p. 312). In such a broad framework, *'the different forms of economy are equal members of a higher plurality'* (ibid., p. 341, italics original).

As Elizabeth Nemeth pointed out, one of the things Neurath relied on in tackling this issue was Mach's 'historical-critical' methods, which integrate historical considerations and epistemological analysis (Nemeth 2007, 2013).<sup>12</sup> In his *Mechanics* (1883), Mach argued that analysing how mechanical phenomena are recognized, that is, historically and critically analysing 'how the principles of mechanics have been ascertained, from what sources they take their origins, and how far they can be regarded as permanent acquisitions', was the most effective way to approach 'the gist and kernel of mechanical ideas' (and finally break down the dogma of a mechanical view on nature) (Mach [1883]1960, p. xxii). The metaphysics is nothing other than the concept of having forgotten how something was reached. *Mechanics* thus returned to texts on classical mechanics, such as those by Galilei, Descartes, Kepler, and Newton, and elucidated the history of the main concepts of mechanics (mass, force, absolute space and time etc.) to expose these were no more than artificially constructed abstractions. The early economic research conducted by Neurath revealed traces of conscious use of Mach's methods in overcoming the division between theory and history within economics. In other words, critically analysing how was the dominant view of 'the economy' historically formed, and what was preserved or lost from amorphous reality by the conceptual structure of economics formed as such, was one way for breaking down the apriorism lurking within economic science and 'freeing up one's point of view.' History had a crucial role in Neurath's attempt to reconstruct economics.<sup>13</sup>

In fact, 'On the Theory of Social Science' (1910) can be seen as showing Neurath's position on *Methodenstreit*. While retrospectively analysing the history of economics, he criticized the tendency of contemporary economists of trying to reduce the subject of 'political economy' to the 'science of exchange', or the 'catalactics' (1910, p. 270). Just as Mach attempted to dismantle unchangeable entities such as the 'atom' and 'molecule' of Newton, or mechanics concepts such as absolute space and time, Neurath criticized, as metaphysical ideas created by extreme analogies with mechanics, the artificial and abstract economic motive of 'self-interest', the picture of the human agent granted such motive as *homo economicus*, and the marginal utility theory as a given premise.

Mechanical analogies in particular propelled the science of *homo economicus*. It was thought that in order to be able to deduce anything one had to apply forces to an object. It goes without saying that this object had to be very simple and completely isolated—that one could put such things to good use was known from mechanics. Thus *homo economicus* was created, as Pareto remarked, an abstraction just like *homo religious*, or *homo ethicus*. It is never mentioned that

<sup>12</sup>Turk (2016) rather stressed the influence of Poincaré's critique of scientific language on Neurath's methodology of economics.

<sup>13</sup>Poettinger (2016) investigates the relationship between history and theory in Neurath's economics methodology. She argues that, in Neurath's view, '[t]he economists...had to be a sociological technician, with a profound knowledge of the past, who was able to extract from the historical discourse possible organizational models, and who possessed a vivid imagination, necessary to be build fictional utopias' (2016, p. 81).

the latter types were never put to any use and that they are useless for psychological analysis. This *homo economicus* then was assigned as the object of his activity something which the old political economy of the mercantilists had not yet known, something that has not been defined adequately even today: the economic good. In order for him to love and desire this object, the *homo economicus* was endowed also with an economic motive that until then had been alien to psychology and will probably remain so in the future. ...It became necessary to create a simplified science of motivations; partly under this influence, marginal utility theory was developed, a purely psychological discipline. (1910, pp. 273–274)

Neurath argued that, as *homo economicus* and the concept of rationality based on monetary calculations were linked, there was a tendency to narrow the target of economic science down to the theory of market relations and price formation, while eliminating extensive ‘change in wealth’ in non-market conditions. For Neurath, the entire conceptual structure of economics was exclusively adapted to the market economy, and building concepts such as ‘capital’, ‘commodity’, ‘value’, and ‘national income’ was determined by their connection with money (cf. Neurath 1920). Therefore, to broaden our theoretical horizon, the structure of concepts had to be significantly reformed.

### 3.2 *Felicitology as Multidimensional Analysis of Human Welfare*

After ‘Political Economy and Value’, in 1911, to reconstruct more generally a theoretical framework that was not limited to the rationality of *homo economicus* with his peculiar economic motives, he adopted ‘wealth’ as the analysis subject of his economic science, instead of ‘relations of exchange and their laws of movement.’ Neurath argued that the concept of wealth, which had roots in Aristotle’s *Nicomachean Ethics*, was a category that rapidly declined as political economy transformed into catallactics.

We find the oldest origins of political economy (*Nationalökonomie*) in the science of household economics on the one hand and in the science of government on the other. The economics of free exchange has only become the object of examination at a relatively late stage. The object of theory and practice was wealth, where wealth was understood as real income in the widest sense. The question of how a people, how humanity can become happy and rich has stood in the centre of attention of the economic literature for a long time. For Adam Smith real income still plays a decisive role. Occasionally he tried to establish the connection between certain economic orders and wealth. His followers have gradually chosen as main object of inquiry the order of monetary and credit relations, which he dealt with in detail, and let fade into the background entirely the question of how the different possible economic orders impinge on wealth...a consistent analysis of change in wealth has not generally been undertaken. (1916, pp. 300–301)

Consequently, the question is what is this ‘wealth’ as the subject of analysis in economics? Neurath first defined it as ‘the totality of pleasure and displeasure that we find with individuals and groups of individuals’ (Neurath 1911, p. 53; cf. 1916, 1917). Subsequently, he adopted terminologies such as ‘life mood’ (*Lebensstimmung*)

or 'happiness'. Life mood refers to the totality of feelings linked with all types of experiences, such as 'eating, drinking, reading, artistic sensibility, religious sights, moral reflection, love, hate, brave and cowardly behaviour' (1917, p. 313). It is a wider concept than the utility used by economists, and by using it instead of utility, he was trying to avoid the premise that humans rationally pursue their own happiness and pleasure, based solely on special economic incentives. Any pleasure and pain that form the life mood are qualitatively unique: the pleasure of eating is one thing, and the pleasure of enjoying art or the pain associated with hatred and anger are different. Additionally, two aspects of a life mood can be pointed out.

First, Neurath clearly positions himself in the Epicurean tradition, by primarily characterizing human welfare in terms of subjective mental conditions. Invoking the words of Epicurus, he affirms:

*Men here on earth who flee sorrow and pain and wish to be kind to each other, happiness, friendship, life as it is really lived on earth, these are our concern; Speculation concerns us only so far as it helps to shape life and to make it happy: "Philosophy is activity in thought and speech, which produces a happy life." "Vain is the ward of a philosopher which does not heal any suffering of man."* (1928, p. 288, italics original)

Unlike Plato, Epicurus admitted women and slaves to the practice of philosophy, and placed no weight on high intellectual achievements. For Neurath, an important issue in economics was also to ask, 'not how can men be purified towards the sublime, but how must you men order your lives to be happy, such as you are' (ibid., p. 288). The following is the reason why he called his economics felicitation. The willingness or auxiliary motive supporting the Neurath's investigation of economics was, in short, an interest in making possible a happy life for 'man on earth'—a concept found in Epicurus and the social philosophy of social reformer Josef Popper-Lynkeus, whom he embraced.

Second, life mood or happiness itself cannot be direct subjects of empirical economics. Because it is a purely private subjective feeling, and even if reasoning by empathy is done to some extent on a daily basis, this is inherently not testable and impossible to measure from the outside. In his 1912 lecture, 'The Problem of a Pleasure Maximum', similar to Robbins' subsequent actions, Neurath argued there were no means of measuring 'degrees of pleasure' and comparing them in size on a cardinal scale, even by reference to the same type of 'pleasure' and the same individual. Therefore, he concluded that 'since until today we have not succeeded in calculating pleasure sums of groups of persons under all circumstances, nor even to state how such a calculation is to proceed, the principle of maximum happiness, as our critique has shown, can never be the basis of a moral or legal system or of a whole order of life' (1912, p. 119). However, Neurath took a totally different direction from the welfare economics' after Robbins: he did not completely abandon welfare comparison between individuals and groups. Instead of directly grasping the life mood, he focused on the objective, publicly observable materials, and social conditions that make it possible to pursue personal happiness (i.e. the perspective of 'physicalism'). He called it 'the inner condition' of happiness, or 'life condition' (*Lebenslage*).

We can, for example, state what food the individuals consume per year, what their housing conditions are, what and how much they read, what their experiences are in family life, how much they work, how often and how seriously they fall ill, how much time they spend walking, attending religious services, enjoying art, etc. (1917, p. 326)

Living standard is the central concept of all circumstances that condition—comparatively directly—a man’s mode of behavior, his sorrows and joys. Shelter, food, clothing, health, books, theatres, friendly human surroundings, all this belongs to the standard of living, even the quantity of malarial germs that menace him. (1931, p. 401)

As a result, the list of life conditions includes not only material or physical conditions, such as shelter, food, and morbidity, but also factors related to the quality of social relations and social institutions in a broad sense, such as working and leisure time, arts and education, freedom and self-government. Neurath emphasized the importance of understanding life conditions as essentially ‘multidimensional structures’ that cannot be evaluated appropriately by a single monetary measure, such as consumption and income (1937, p. 146). Particularly, the flaw of traditional economics is it failed to grasp the impact on welfare caused by factors that cannot be subjects of ‘demand’, such as poor working environment, infant mortality rate, and hostile social relations (ibid., p. 141). Therefore, Neurath argued that felicitology should transgress the bounds of analysis on pure market relations.<sup>14</sup>

Interestingly, Neurath grasped life mood and life conditions in conjunction with ‘external condition’ of happiness, that is ‘basis of life’ (*Lebensboden*). The basis of life is the ‘environment in the broadest sense’, which includes sources of all raw materials, various energy sources, minerals, animals and plants, micro-organisms, soils, forests, rivers, climatic conditions, and so on. Neurath also attempted to understand the ‘order of life’ (*Lebensordnung*), that is, institutions and types of human behaviour, as ‘embedded’ in biological and geographical conditions, and as the structures that interact with them: ‘Society appears as a structure moved by the conditions of its environment to which its own parts also belong’ (Neurath 1931, p. 393). Neurath’s viewpoint on the environmental conditions of human life are also found in the discussion on socialization from the 1920s. With reference to the history of modifying ecological conditions by human economic activities, such as a change of vegetation and the relationship between deforestation and climate change, he pointed out that ‘the most difficult problem’ was one ‘of the exploitation of a given basis of life by the economic order’. Later, this issue had a significant meaning in his concept of economic rationality and critique of capitalistic economic order (1920, p. 349; cf. 1931, pp. 395–394).<sup>15</sup>

---

<sup>14</sup>Regarding life condition, its similarity with Amartya Sen’s capability approach has been noted. Leßmann (2007) summarizes their similarity on the issue (1) of advancing the idea of a new concept tool to evaluate human welfare from the measurability of utility and individual comparison, and (2) of criticizing the informational base, such as income and utility, or emphasizing the plurality of different components constitutive human welfare. See also O’Neill and Uebel (2008).

<sup>15</sup>At this point, Neurath’s economic writings were rediscovered as precursors of ecological economics. See Martinez-Alier (1987), Uebel (2005b), and O’Neill and Uebel (2015).

As background to Neurath's discussion, the influence of the geographic tradition and geopolitics in Germany, such as Friedrich Ratzel's magnum opus *Anthropogeographie* (1882–1891) is evident. Additionally, in the *Foundations of Social Sciences*, published in 1944 in the series *Foundations of the Unity of Science*, he cited from British human geography and human ecology studies from the early twentieth century,<sup>16</sup> and identified the analysis subject of social science as one field of unified science, using the ecological term 'synusia' or 'cosmic aggregation': 'We suggest not starting with the antithesis: living being and the environment (as bioecology does), but starting with what may be called a "synusia" composed of men, animals, plants, soil, atmosphere, etc. I am using the term "synusia" in analogy with the term "symbiosis"' (1944, p.20). Here, human subjects are also not rational *homo economicus*, abstracted from social interdependence, but are understood as 'a geological agent' in a complex entanglement with various physical wealth, resources, and bio-geographical conditions.

In this way, Neurath conceived an economic science (felicitation) as an empirical one, based on in-kind statistics, to investigate how the complex correlations between different economic orders and objective natural and social conditions change the subjective experiences of individuals:<sup>17</sup>

Social scientists may characterize 'living conditions' by telling of shelter, food, entertainment, friendship, 'oceanic feeling', or whatever one thinks of importance. Some of these items may be measured in their own units (e.g., calories, cubic space used in houses, etc.), but in other cases we get only the degrees of certain items which all together constitute what we may call a 'silhouette' of living conditions. ...Such and similar 'felicitation' investigations are properly empiricist, and we may prepare sociological questionnaires which deal with these items. We may start from the same 'basis of life' (i.e., an aggregation composed of human being, animals, swamps, fields, houses, micro-organisms, etc.) and ask what different living conditions appear by assembling the same items in a different way. (Neurath 1944, p. 37)

Felicitation thus focused on natural and social conditions that provide the possibility of pursuing happiness for 'men on earth', not the utility of individuals or subjective pleasures. Neurath called this position 'social Epicureanism', which is distinct from utilitarianism. In doing so, similar to Robbins, he frequently emphasized that felicitation belongs to theoretical economics, which is specialized in the description of facts, not practical economics: 'Making happiness the subject of special investigations does not mean that people act exclusively for the sake of happiness nor that they should do so; nor does it mean that happiness plays a particularly important

<sup>16</sup>These include R. L. Sherlock's *Man as Geological Agent* (1922), Charles Elton's *Animal Ecology* (1927) and J. W. Bews' *Human Ecology* (1935).

<sup>17</sup>Felicitation needed a way to describe and represent the elements that make the lives of human beings and their interrelations different from 'geometric representations' such as Edgeworth's box diagram. ISOTYPE (International System of Typographic Picture Education), which uses a visualization method to express a social state consisting of heterogeneous elements, such as resource/energy consumption, working time and leisure time, occupation conditions, birth rate and death rate, was an attempt to depict the 'topography of life conditions' of social groups for individuals (including workers without literacy skills) to understand it. See also Leonard (1999) and Nemeth (2013).

role in the structure of the world. It is only dealt with as a fact, as a May-bug is dealt with a zoologist without claiming that May-bugs are something especially important, or that May-bugdom has to be promoted everywhere' (1917, p. 315). Felicitation, as empirical science, was an attempt to restore a wider perspective to economics, which can describe the relationships of wealth that cannot be represented in monetary terms, and also criticizes contemporary liberal economics falling into the fallacy of 'the application of augments, valid in the field of higher bookkeeping, to the analysis of social problems and human happiness in general' (1944: 39).

However, especially during the era of Red Vienna, his felicitation had meaning not only as an empirical science, but also as an intellectual tool for the conscious and democratic formation of the various conditions of social life, that is, 'socialization' in the near future. Neurath was confident that his renewing of economic science would 'enable us to significantly enlarge our theoretical horizon', thereby 'seeds of future developments could be recognized as such sooner than before and the significance of many events would be better appreciated' (1916, p. 308). However, for Neurath, science is not just a neutral instrument that should serve social and political ends. While the values associated with these ends remain outside science, the choices they inform help to shape the direction and content of science (Okruhlik 2009, p.61). In sum, felicitation was an essential intellectual program to enabling socialization, being supported, at the same time, by his own auxiliary motive, 'the will to socialization'.

#### **4 Will to Socialization, or Rationality Without Foundations**

After the First World War, Neurath theoretically and practically explored the socialized economy, based on calculations *in natura*, not using money. This idea of economic planning is relatively well known because of criticisms by Weber, Mises, and Hayek. For Neurath, the decisive significance of felicitation on socialization after the war was that, whereas every economic activity is organized by the purpose of maximizing 'profit' under capitalistic order, maximizing 'happiness' for all individuals is the purpose of socialism. He thus emphasized the fundamentally different measures of rationality in capitalism and socialism: 'The goal of individual enterprises in the capitalist economy is the maximum gain of money and because of this money calculation makes sense namely, to establish whether the maximum has been reached. In a socialist economy the goal is the maximum of happiness and quality of life for everyone, of utility, and because of this the calculation of utility, happiness and quality of life make sense. (As far as we know this is impossible to do in units)' (1925, pp. 471–472). One rationality is 'profitability' (formal rationality), aimed at maximizing money, while the other rationality is maximizing the wealth (happiness) of all individuals living in a society, that is, 'the productivity of economy' (substantive rationality). Based on this distinction, Neurath pointed out that categories themselves, such as profit and cost, have no meaning in the *in natura* framework calculations, which is the essential basis of the socialist calculation of economic rationality:



'Costs' in the capitalistic sense and the 'negative quantities' of socialist calculation do not come to the same thing.... The 'positive quantities' of the socialist order also do not come to the same thing as the 'profit' of capitalism. Savings in coal, tree, etc., beyond amounting to savings in the displeasure of work, mean the preservation of future pleasure, a positive quantity. For instance, that coal is used nowadays for silly things is to be blamed for people freezing in the future.... *The future figures in the balance sheets of the capitalistic order only in so far as the demand is anticipated.* The freezing people of the future only show up if there is already now a demand for future coal. Just as before, capitalism cuts down the forests even if the consequence may be karstification in a hundred years. In the tropics, and elsewhere, capitalism engages in over-exploitation without any care. In short, for capitalism such savings would mean a loss of profit. (1925, pp. 470–471, italics added)

As pointed out in the second half of the above quote, the pursuit of economic rationality based on market prices can be just 'the irrational rationality', in that it might have a destructive impact on the actual life conditions of individuals. This is because '[t]he money calculation of the economic order of capitalism is very precise in terms of money sums, but it tells us nothing about the true "wealth" of a people, neither about the use made of sources of raw materials nor about the distribution of the good produced' (ibid., p. 468). For Neurath, mass poverty and unemployment, associated with periodic crises, quick changes in fashions inspired by the advertising industry, the urgent needs of the poor being ignored because of their lack of purchasing power, premature resource depletion, waste of all types of energy, and the destruction of real physical preconditions for the life of future generations, were all evidence that, although the monetary calculations in the market system allow for advanced economic calculability, they also created a serious temporal myopia of the economic agents and irrational results for the substantive life on the earthly plane.<sup>18</sup> Neurath understood that Mises' thesis of 'without calculation with *one* unit, an economy is *not* passible' was just the fallacy of pseudorationalism, which treats everything as calculable, analysable, and representable by precise formulas (1925, p. 430).

By contrast, in a socialized economy aiming to maximize happiness, there is no means to calculate the rationality of planning based on a common denominator. As such, Neurath's calculation *in natura* had nothing to do with algorithmic calculations. Rather, it is related to the fuller recognition of various different material and social conditions of the lives of individuals—including not only the current generation, but also potential future ones—which are mutually non-reducible. To formulate a rational economic plan, '[t]he most careful and conscientious, perhaps statistical, consideration of all circumstances does not give us a unit for calculation' (1928, p. 263). For the planning to be rational, an overall unified knowledge is necessary, for 'the stock of all connectable and indeed logically compatible laws' (1931, p. 329) to describe and predict real human life conditions. For example, in agricultural economics, we have to mobilize all of relevant knowledge about 'how the same fields, woods, human and animal forces can produce various components of conditions of life by applying different technical, biological, chemical, etc., methods' (1917, p. 318). In the other

---

<sup>18</sup>Neurath's critique on the view of economic rationality based on market price, as well as Polanyi's, had an important meaning for K. W. Kapp's conception of 'social costs'. See Berger (2015).

words, unified science as ‘the orchestration of the sciences’ was, above all, an epistemological condition of the practice of comprehensive control of ‘the economy’, which was composed of a great variety of distinctive things.

However, even if one succeeds in forming such unified knowledge, it does not mean a specific plan would be uniquely guided by it, or would ever result in a rigorous foundation that guarantees the rationality of this plan. Neurath emphasized there exist no calculation units, equations, or technocratic solutions that will lead to ‘the one best solution’ to social-economic problems with ‘optimum happiness’, ‘optimum population’, ‘optimum health’, ‘optimum working week’, and ‘optimum productivity’. Even if scientific knowledge contributes to the qualitative improvement of democratic planning, as far as ‘the decisions are different when made without comprehensive knowledge, or when made after hearing all the experts’, ‘if they were asked to make decisions and not only to prepare arrays of possible solutions it’, it will ‘even disturb the scientific habit of experts’ (1942, pp. 426–427). There are two reasons for this.

As previously discussed, the first is because the inner limit of scientific knowledge cannot escape uncertainty and unpredictability, and the second is emphasized in his 1942 essay, ‘International Planning for Freedom’, being related to the nature of human welfare. According to Neurath, human welfare is associated with ‘custom’ and ‘tradition’, as it is often incompatible with the ‘technical efficiency’ presented by experts or scientists. Therefore, social engineers must realize more than such efficiency, for instance, that ‘all homely comfort relates to certain traditional customs and environments and that joy sometimes might depend solely on the fact that something should *not* be changed according to certain technical and architectural standards’ (ibid., p. 423).

Assume the scientists tell the English people that their fireplaces waste calories—of course they do so enormously. But the fireplaces as an element of our environment are not ‘happiness-neutral’ as it were, as is e.g., the cable shaft below the surface of the street. The fireplaces are related to homely comfort and to many customs of our private life. How to compute these and other items of ‘happiness conditions’ would be the subject of discussions and finally of decisions based on common sense and influenced by the scientists’ information. (ibid., 427)

All decisions and judgements on planning and socialization could become justifiable only by showing they fit into the patterns of personal life of which we approve. In sum, the rationality of socialization and planning is *rationality without foundations*, and depends on democratic institutions in a broader sense, which pursue the goal of ‘orchestration of ways of living’. Neurath explicitly criticized totalitarian absolutism by arguing:

People of the totalitarian kind may try to make scientists the leaders of a new society, like the magicians, nobles, or churchmen of former societies. The encyclopedism of logical empiricism does not see why scientists, trained to discover as many alternatives as possible, should be particularly able to select one alternative only (one that never can be based on calculation) by making a decision or performing an action for other people with different desires and attitudes. (1946, p. 505)

Moreover, Neurath emphasized that social engineering should consider that the planning instruments or forms themselves were also inextricably related to their

scopes. This is because 'personal independence', 'voluntary cooperation', 'democracy', and 'self-government and freedom', in opposition to 'rigid order', 'superimposed regulations', and 'one-party system', are all essential elements of the 'happiness conditions' of human beings (1942, p. 423). As a result, social engineers should test a social order and its institutions not by its 'business efficiency,' but its ability to produce happiness conditions such as food, shelter, education health, and freedom. Neurath maintained that the freedom of a democratic society was best 'produced' within a pluralistic socio-economic system (he called it '*societas societatum*'), where 'each member is permitted to have more than one loyalty, e.g. to his family, to his local community, to his lodge, to an international movement and to his country' (ibid., 429).

Indeed, this last theme on the interdependence of freedom and democratic planning had something to do with the liberal critiques of socialism, including Hayek's.<sup>19</sup> In the review of *The Road to Serfdom* in 1944, Neurath agreed with Hayek's argument that 'there are people who like planning as a new means of obtaining totalitarian leadership and there are others who unwittingly support fascism by promoting certain principles of planning' (1945, p. 546). On the other hand, Neurath strongly objected to Hayek's dualism that set the alternative of 'the freedom of market competition' or '*the planner's* unlimited totalitarianism', and his diagnosis that all forms of planning would be 'the road to serfdom'. Rather, the blind praise of market competition and 'the internationalism of the "money order"' were contradicting the pluralism of logical empiricists, who fought any 'pseudo-rationalism' that regards one solution as the only and best, and found that a multitude of further possibilities and intermediate solutions always lay open (cf. Zolo 1989, p. 162). For Neurath, 'planning as a co-operative effort, based on compromise' could give rise to 'a world-wide feeling of responsibility for other people's happiness', which Hayek deliberately ignored, and be compatible—not necessarily so—with the realisation of the values of 'freedom' and 'democracy' as 'the acknowledgement of "non-conformism" within a freely accepted social order' (1945, pp. 546–547).

## 5 Conclusion

Neurath's felicitology could show a potential of empirical (physicalist) economics far from the conventional interpretation of the relationship between logical empiricism and social sciences in the early twentieth century. Neurath transformed the picture of the economic agent from *homo economicus*, which does not have a counterpart in empirical reality, to the 'man on earth' or 'man as geological agent', embedded and embodied within the social and ecological relations of interdependence. Moreover, he attempted to reconstruct the conceptual structure of economics as a science that comprehends and predicts concrete living and happiness conditions of such agents.

---

<sup>19</sup>On Neurath's non-market associational models of socialism and his later debates with Hayek, see O'Neill (2003, 2006).

This was clearly not a science studying rational behaviour, but human happiness. In the early twentieth century, it could be identified as alternative economic thought to marginal utility theory, developed as a purely psychological discipline based on a peculiar economic motive, that is ‘self-interest’. Neurath’s felicitology would then be understood as one origin of the genealogy of ‘substantive economics’ of Karl Polanyi and K. W. Kapp, in that it not only criticized the limits of the market’s purely formal rationality in terms of the structure of existential human needs, but also considered the complex interdependence between freedom and social planning, against Mises’ and Hayek’s critiques on the idea of planning.

## References

- Berger, S. 2015. K. William Kapp’s *Social* theory of social costs. *History of Political Economy* 47 (annual suppl.): 227–252.
- Carnap, R., H. Hahn, and O. Neurath. 1929. Scientific world conception. The Vienna circle. In *Empiricism and sociology*, ed. R.S. Cohen and T. Uebel, 299–318. Dordrecht: Reidle.
- Carnap, R. 1963. Carnap’s intellectual autobiography. In *The philosophy of Rudolf Carnap*, ed. P.A. Schilpp, 3–84. LaSalle, IL: Open Court.
- Cartwright, N., J. Cat, L. Freck, and T. Uebel. 1996. *Otto Neurath: Philosophy between science and politics*. Cambridge: Cambridge University Press.
- Cat, J., N. Cartwright, and H. Chang. 1996. Otto Neurath: Politics and the unity of science. In *The disunity of science*, ed. P. Galison and D.J. Stump, 347–369. Stanford: Stanford University Press.
- Hands, D. Wade. 2001. *Reflection without rules: Economic methodology and contemporary science theory*. Cambridge: Cambridge University Press.
- Hands, D. Wade. 2010. Economics, psychology and history of consumer choice theory. *Cambridge Journal of Economics* 34: 633–648.
- Haller, R. 1988. *Questions on Wittgenstein*. London: Routledge.
- Haller, R. 1991. The Neurath principle: Its ground and consequences. In *Rediscovering the Forgotten Vienna Circle*, ed. T. Uebel, 117–130. Dordrecht: Springer.
- Hayek, F.A. 1952. *The counter-revolution of science: Studies on the abuse of reason*. London: The Free Press.
- Hollis, M., and E.J. Nell. 1975. *Rational economic man: A philosophical critique of neo-classical economics*. Cambridge: Cambridge University Press.
- Leßmann, O. 2007. A similar line of thought in Neurath and Sen: Interpersonal comparability. In *Otto Neurath’s economics in context*, ed. E. Nemeth et al., 115–130. Dordrecht: Springer.
- Leonard, R.J. 1999. ‘Seeing is believing’: Otto Neurath, graphic art, and the social order. *History of Political Economy* 31 (supplement): 452–478.
- Mach, E. [1883]1960. *The science of mechanics: Critical and historical accounts its development*, 6<sup>th</sup> edn, transl by T. J. McCormack, Illinois: Open Court Publishing.
- Martinez-Alier, J. 1987. *Ecological economics: Energy, environment and society*. Oxford: Blackwell.
- Nemeth, E. 2007. Freeing up one’s point of view: Neurath’s Machian heritage compared with Shumpeter’s. In *Otto Neurath’s economics in context*, ed. E. Nemeth et al., 13–36. Dordrecht: Springer.
- Nemeth, E. 2011. Scientific attitude and picture language: Otto Neurath on visualisation in social sciences. In *Image and imaging in philosophy, science and the arts*, ed. R. Heinrich et al., vol. 2, 59–83. Frankfurt: Ontos Verlag.

- Nemeth, E. 2013. The philosophy of the other Austrian economics. In *New challenges to philosophy of science: The philosophy of science in a European perspective*, ed. H. Andersen et al. vol. 4, 339–350. Dordrecht: Springer.
- Neurath, O. 1910. Zur Theorie der Sozialwissenschaften, trans. as On the theory of social science. In *Otto Neurath: Economic writings: Selections 1904-1945*, ed. T.E. Uebel and R.S. Cohen, 2004, 265–291. Dordrecht: Kluwer.
- Neurath, O. 1911. Nationalökonomie und Wertlehre, eine systematische Untersuchung. *Zeitschrift für Volkswirtschaft, Sozialpolitik und Verwaltung*, Bd. 20, 52–114.
- Neurath, O. 1912. Das Problem des Lustmaximums, trans. as The problem of the pleasure maximum. In *Empiricism and sociology*, ed. M. Neurath and R.S. Cohen, 1973, 113–122. Dordrecht: Reidel.
- Neurath, O. 1913. Die Verirrten des Cartesius und das Auxiliarmotiv, trans. as The Lost wanderers of Descartes and the auxiliary motive. In *Otto Neurath: Philosophical papers*, ed. R.S. Cohen and M. Neurath, 1983, 1–12. Dordrecht: Reidal.
- Neurath, O. 1916. Die Naturalwirtschaftslehre und der Naturalkalkül in ihren Beziehungen zur Kriegswirtschaftslehre, trans. as Economics in kind, calculation in kind and their relation to war economics. In *Otto Neurath: Economic writings: Selections 1904-1945*, T.E. Uebel and R.S. Cohen, 2004, 299–311. Dordrecht: Kluwer.
- Neurath, O. 1917. Das Begriffsgebäude der Wirtschaftslehre und seine Grundlagen, trans. as The conceptual structure of economic theory and its foundations. In *Otto Neurath: Economic writings: Selections 1904-1945*, ed. T.E. Uebel and R.S. Cohen, 2004, 312–341. Dordrecht: Kluwer.
- Neurath, O. 1920. *Vollsozialisierung*, trans. as Total socialization. In *Otto Neurath: Economic writings: Selections 1904-1945*, ed. T.E. Uebel and R.S. Cohen, 2004, 371–404. Dordrecht: Kluwer.
- Neurath, O. 1925a. *Wirtshtsplan und Naturalrechnung*, trans. as Economic plan and calculation in kind: On the socialist order of life and the human beings of the future. In *Otto Neurath: Economic writings: Selections 1904-1945*, ed. T.E. Uebel and R.S. Cohen, 2004, 405–465. Dordrecht: Kluwer.
- Neurath, O. 1925b. Sozialistische Nützlichkeitsrechnung und kapitalistische Reingewinnrechnung, trans. as Socialist Utility Calculation and Capitalist Profit Calculation. In *Otto Neurath: Economic Writings: Selections 1904-1945*, ed. T.E. Uebel and R.S. Cohen, 2004, 466–472. Dordrecht: Kluwer.
- Neurath, O. 1928. *Lebensgestaltung und Klassenkampf*, trans. as Personal life and class struggle. In *Empiricism and sociology*, ed. M. Neurath and R.S. Cohen, 1973, 249–298. Dordrecht: Reidle.
- Neurath, O. 1931. *Empirische Soziologie*, trans. as Empirical sociology: The scientific content of history and political economy. In *Empiricism and sociology*, ed. M. Neurath and R.S. Cohen, 1973, 319–421. Dordrecht: Reidel
- Neurath, O. 1932. Protokollsätze, trans. as Protocol Statements. In *Otto Neurath: Philosophical papers*, ed. R.S. Cohen and M. Neurath, 1983, 91–99. Dordrecht: Reidal.
- Neurath, O. 1937. Inventory of the standard of living, *Zeitschrift für Sozialforschung* 6, 140–151.
- Neurath, O. 1942. International planning for freedom, repr. In *Empiricism and sociology*, M. Neurath and R.S. Cohen, 1973, 422–440. Dordrecht: Reidle.
- Neurath, O. 1944. Foundations of social sciences. In *Foundations of the unity of sciences: Towards an international encyclopedia of unified science*, ed. O. Neurath, R. Carnap, C. Morris, vol. 2, 1970, 1–51. Chicago: University of Chicago Press.
- Neurath, O. 1945. Alternatives to market competition (Review of F. Hayek *The Road to Serfdom*), repr. In *Otto Neurath: Economic writings: Selections 1904-1945*, ed. T.E. Uebel and R.S. Cohen, 2004, 546–548. Dordrecht: Kluwer.
- Neurath, O. 1946. The Orchestration of the sciences by the encyclopedism of logical empiricism. *Philosophy and Phenomenological Society* 6 (4): 496–505.
- Neurath, O. 1973. *Empiricism and sociology*, trans. and ed. M. Neurath, and R.S. Cohen. Dordrecht: Reidle.
- Neurath, O. 1983. *Otto Neurath: Philosophical Papers 1913–1946*, trans. and ed. R.S. Cohen, and M. Neurath. Dordrech: Reidel.

- Neurath, O. 2004. *Economic writings: Selections, 1904–1945*, trans. and ed. Uebel, T.E., and R.S. Cohen, Translated by R.S. Cohen, M. Neurath, C. Schmit-Petri, and T.E. Uebel, 546–548. Dordrecht: Kluwer.
- Okruhlik, K. 2004. Logical empiricism, feminism, and Neurath's auxiliary motive. *Hypatia* 19 (1): 48–72.
- O'Neill, J. 2003. Socialism, associations and the market. *Economy and Society* 32 (2): 184–206.
- O'Neill, J. 2006. Knowledge, planning, and markets: A missing chapter in the socialist calculation debates. *Economics and Philosophy* 22 (1): 55–78.
- O'Neill, J., and T. Uebel. 2008. Logical Empiricism as Critical Theory? The Debate Continues. *Analyse & Kritik* 30: 379–398.
- O'Neill, J., and T. Uebel. 2015. Analytical philosophy and ecological economics. In *Handbook of ecological economics*, ed. J. Martinez-Alier and R. Muradian. Cheltenham: Edward Elgar.
- Poettinger, M. 2016. Between history and theory: Otto Neurath's economics from 1906 to 1917. In *Economic thought and history: An unresolved relationship*, ed. Poettinger and Tusset, 75–99. London: Routledge.
- Putnam, H. 2002. *The collapse of the Fact/Value dichotomy*. Harvard: Harvard University Press.
- Robbins, L. 1935. *On the nature and significance of economic science, (first published, 1932)*, 2nd ed. London: Macmillan.
- Sandner, G. 2014. Political polyphony: Otto Neurath and politics reconsidered. In *European philosophy of science: Philosophy of science in Europe and the Viennese Heritage*, ed. Galavotti, M.C. et al., 211–222. Dordrecht: Springer.
- Sen, A.K. 1987. *On ethics and economics*. Oxford: Blackwell.
- Stuchlik, J. 2011. Felicitology: Neurath's naturalization of ethics. *The Journal of the International Society for the History of Philosophy of Science* 1 (2): 183–208.
- Turk, M. 2016. Otto Neurath and the linguistic turn in economics. *Journal of the History of Economic Thought* 38 (3): 371–389.
- Uebel, T. 2005a. Political philosophy of science in logical empiricism: The left Vienna circle. *Studies in History and Philosophy of Science* 36 (4): 754–773.
- Uebel, T. 2005b. Incommensurability, ecology, and planning: Neurath in the socialist calculation debate, 1919–1928. *History of Political Economy* 37 (2): 309–342.
- Walsh, V. 1987. Philosophy and economics. In *The New Palgrave: A dictionary of economics*, vol. 3, ed. J. Eatwell, M. Milgate, and P. Newman, 861–869. London: Macmillan.
- Walsh, V. 1996. *Rationality, allocation and reproduction*. Oxford: Oxford University Press.
- Zolo, D. 1989. *Reflexive epistemology: The philosophical legacy of Otto Neurath*. Dordrecht: Kluwer Academic Publishers.

# Selfish, Therefore Reciprocal: The Second Marginal Revolution of Mises



Akihiko Murai

**Abstract** Selfishness in economics seems to appear typically in utility-valuation theory of consumer. But it is not the case. While British moral philosophers sought to mitigate it with conscience, no integration was achieved. In the Marginal Revolution, the *unconscious* and *passive* acceptance of saturation kept the selfish agent from realizing his aim, as he made to consume undefined amount of goods at whose *last unit* he stops consuming for the advent of saturation, where Gossen's second law of equal marginal utility holds. Only Menger, however, denied such an exchange. Departing from his cardinal guise, Mises improved the theory dissociating selfishness from saturation, which a rational agent should deny. Rothbard formulated this alternative as the *relevant unit* of defined amount of goods. When one makes exchange according to this principle, selfishness will bring reciprocity. Thus, long-awaited integration of selfishness with societal benefit is established. Morgenstern, on the other hand, took a path to re-cardinalize the theory relying on the equivalent exchange scheme and substituted utility of money for that of goods. But the validity of his Expected Utility Theory is dubious and the Misesian purely ordinal theory seems quite justifiable even now.

**Keywords** Theoretical historiography · Hobbesian paradigm · Praxeology · Relevant-unit marginality · Saturationism · Bird-eye assessment

## 1 Introduction

This chapter is an essay regarding the problem of selfishness<sup>1</sup> in terms of the theory of utility in relation to exchange. History requires a goal, or at least conviction in a certain improvement toward it. But we do not have the *general theory of utility*

---

<sup>1</sup>We find this term as well as “self-love” or “self-interest” in various literature. In this chapter, however, we use “selfishness” as a noun unless the original writer expresses otherwise, for distinction in relation to economic analysis, not ethical, is not important within the framework here.

---

A. Murai (✉)  
Kansai University, Osaka, Japan  
e-mail: [a.murai@aist.go.jp](mailto:a.murai@aist.go.jp)

applicable to any type of goods. Some may say that the Expected Utility Theory provided it. However, as we see below, Morgenstern himself refused its completion. While any history of theory should depict the inner development process, we have no goal in textbook economics. Economists sometimes refer to preceding theories, but only some aspects of the development are described. Utility theory dates back to Aristotle, and economists seem to address the same problem from relatively close perspectives. But when there is no goal, its history remains an interim report of various unsuccessful endeavors. If, however, we look to another paradigm, we may find a more integrated alternative. Thus, we are obliged to reconsider the endeavors from an *inter-paradigmatic* perspective.

This analytical perspective can be characterized as a *theoretical historiography*, in which theories are broken down to unit ideas, or essential components, as was suggested by Lovejoy (1936: 3), and a common platform is assumed to attain a certain interface between not only theories but even paradigms. For arguments about utility are related to the valuation theory of a person making an exchange, and selfishness is another name for the desire to acquire some gain from it, this requires us to work out a *comparative catallactics*. Each contributor is seen from the common viewpoint. This does not involve relative ignorance of text, but rather is based on the respective texts themselves. It looks into some neglected messages in these texts: Examination of diverse materials from a fixed viewpoint will make the text show what has escaped attention and yet still arrive in readers' consciousness though written evidently.

The selfish economic man has long been criticized. But finding a truly economizing person in textbook economics is not easy. It will be quite clear that the economic man is less selfish than generally conceived. On the other hand, Mises envisaged different type of selfish person. And in so doing he invented *another* marginal principle, i.e., *there were two marginal principles*. Reviewing the history of theories about selfishness for roughly three centuries, we conclude that Mises caused the *second Marginal Revolution* and that he succeeded in establishing a truly selfish agent in economics within the framework of his *praxeology*, or the science of human action. It is this insight that enables us to see the history of utility theory from a different perspective.

## 2 Selfishness Without Catallactics

In the British context, the underlying idea of economic theorizing in terms of selfish agents seems to have been laid down by Thomas Hobbes (1588–1679). He made a magnificent attempt to create a revolutionary theory of man and society in the form of “psychophysics” (Husserl 1970: 62). Having a pessimistic view on the natural condition of man, he was an ardent advocate of the absolute monarchy. But at the same time he wanted more peace than war (Hobbes 1983: 50), because he knew the primacy of self-preservation. This emphasis on selfishness, however, was mitigated by his view that it was essentially evil. He wrote no book on economics and his economic analyses were dispersed in political or moral philosophical books, which



have an essentially vertical viewpoint. One of his favorite economic subjects was money. But his view was very specific: “Sometimes a man pays a debt only out of fear of imprisonment; but because nobody prevented him from keeping the money, paying it was the action of a man at liberty”(1992: 197). Note the agent is both free and bound here. Thus Hobbes did not discuss economics as horizontal trades between cell citizens of body state. But a long-lasting framework grew out of his perspective: (i) a *necessary-evil view on selfishness* that activities for self-preservation can cause strife, (ii) a *mitigation view of two conflicting factors* to settle or prevent it, and (iii) a *single-person view of selfish valuation*, or confinement of the valuation to from *one party* in exchange. While the last was developed later, we may call this triad a Hobbesian Paradigm.

Hobbes was challenged by the third Earl of Shaftesbury (1671–1713). While well-known for his emphasis on the higher conduct of men, he did not deny selfishness. Rather, he sought a harmonious coexistence of selfishness with public benefit. Although Hobbes saw tension between private and public benefits, they are not only consistent, but inseparable as physical “parts or members” of a creature does not fight against the body but serves as integral parts of it (Shaftesbury 1995: 281–2). But he saw a certain risk in selfishness for it can produce evil if private affection is too strong (247–50).

Shaftesbury was challenged by Bernard Mandeville (1670–1733). The title of his work, *The Fable of the Bees, or Private Vices, Publick Benefits*, implies a negative view on the nature of self-preservation. But even if this is the case in his verse “The Grumbling Hive”, it is not so in the prose. In the latter he presented a class-mentality origin of morality. According to Mandeville, the First Class of mere self-preservers, “vile grov’ling Wretches” or “Dross of their Kind, and having only the shape of men”, paid respect to the Second, “lofty high-spirited Creatures, that free from sordid Selfishness” and conscious about public affairs besides their own self-preservation (Mandeville 1957 [1714]: 43–4). The narrowness of mind was called “VICE” while Second’s bravery or care-taking of the First was called “VIRTUE”(48–9). For even the Second Class care for their own preservation and the problem lies in the *narrowness* of mind rather than self-preservation itself. He also took a moralistic view that vice should be punished in *A Letter to Dion* (2013: 23).<sup>2</sup> Moreover, while abstained from laughing at Shaftesbury’s endeavors (1957 [1732]: 44), he criticized, in support of the revealed religion, the latter’s deism (357).

These arguments made not only the orbit and scope of English debate, but also Scottish. Adam Smith (1723–90), an alleged creator of economics, took up selfishness in detail in *The Theory of Moral Sentiments*. But it could be said that his agent is free and bound again:

---

<sup>2</sup>His view derives from Augustinianism. According to Jacob Viner, if the Second Class man acts out of charity to “endeavour the Benefit of others, or the Conquest of his own Passions out of a Rational Ambition of being good”(Mandeville 1957 [1714]: 48–9), “this definition of virtue is in strict conformity to Augustinian rigorism” and resembles a phrase “*aut caritas aut cupiditas* [either charity or greed]”(2013: 5).

Self-love was a principle which could never be virtuous in any degree or in any direction. It was vicious whenever it obstructed the general good. When it had no other effect than to make the individual take care of his own happiness, it was merely innocent, and though it deserved no praise, neither ought it to incur any blame (Smith 1976 [1759], VII.ii.3.12).

But because he did not delve into the matter in the *Wealth of Nations*, we have little idea of his view on selfishness regarding consumer activities.<sup>3</sup> He saw exchange as a “propensity” of men (1976 [1776]: I.ii.1). From this perspective, consideration of the condition for exchange to be done is useless and the detailed psychological accounts of utility in the *Moral Sentiments* disappeared.

Utilitarians changed the way scholars form their arguments about self-interest. In earlier days, when Christianity had stronger influence on the matter, moralists as well as economists often referred to “conscience” in arguing self-love or self-interest. This is because they believed that an agent could be drawn with a unity of personality in which conscience, or knowledge of the good shared with God, can domesticate selfish drive in various scenes in our life. But the language altered in the nineteenth century. Arguments on conscience disappeared from the subject list of ethics. Instead, an agent taking care of his enjoyment of pleasure and avoidance of pain became a new focus. Jeremy Bentham (1748–1832) saw a pleasure as an axiomatic starting point that needs no proof (Bentham 1962: 1.11–4). Nevertheless, the Hobbesian Paradigm remained intact, for utilitarians paid little attention to the utility in connection with exchange despite their name.

John Stuart Mill (1806–73) inherited the British tradition of finding selfishness “narrow”<sup>4</sup>, and sought to mitigate it not by awakening our conscience, but widening the scope of our brotherhood from the narrower to the broader (Mill 1969: 233). On the other hand, he wrote “On the Definition of Political Economy” in 1836 and it was here that he drew a prototype of the *homo economicus*. In striving for a proper definition of economics, he created an economic man. But this necessitated Mill describing him “solely as a being who desires to possess wealth, and who is capable of judging the comparative efficacy of means for obtaining that end” (1967: 321). In addition, he defined economics as “[t]he science which traces the laws of such of the phenomena of society as arise from the combined operations of mankind for the production of wealth in so far as those phenomena are not modified by the pursuit of any other object” (323). Note that Mill refers to micro aspects of the agent as “operations” and macro results as “phenomena”. In view of this definition, his *homo economicus* could have been a praxeological agent capable of comparing means for an economic aim in terms of efficacy to obtain it, but he did not take this path in his *Principles of Political Economy* about a decade later (1848).

However, a praxeological view of the economic agent was prevalent in the tradition of continental, or, more precisely, Catholic economics. Richard Whatley (1787–1863), a Drummond professor of Oxford University and a professor of Trinity College

<sup>3</sup>His argument on the selfishness of butcher, brewer or the baker and on their unintended public benefit (Smith 1976 [1776], I.ii.2) is well known. But since this opinion is related to the supply-side of the goods market, we exclude it here.

<sup>4</sup>E.g. Hutcheson 1990: 16–7, Godwin 1993: 432, Malthus 1986: 69, and Mill 1977: 305.

Dublin followed this line. Justifying his belated treatment of such a fundamental issue as value theory in the book, Mill referred to “one eminent writer” who suggested the substitution of “catallactics” for “economics”(Mill 1965: 455), which Whatley wrote in his *Introductory Lectures of Political Economy* (Whatley 1832: 6). Mill insisted that his having been able to postpone value theory to Part III was evidence of the fallacy of Whatley. In short, the problem of value can be dealt with independent of micro “operations” theory. Mature Mill seemed to be at odds with younger Mill. Furthermore, he wrote such a bold statement as “Happily, there is nothing in the laws of Value which remains for the present or any future writer to clear up; the theory of the subject is complete”(Mill 1965, 456). It was this passage that Jevons quoted to criticize Mill in his *Theory of Political Economy* in 1871 (Jevons 1957: v, 76).

### 3 Comparative Catallactics of the Marginal Revolution

Thus, the British argument over fundamental issues called for the next economics starting from micro agent’s subjective examination of *utility*. This time the frame of reference was not ethics but catallactics. As conscience lost its role in ethics, so did the integral agent. But this may seem strange, for economic activity itself inevitably presupposes the agent. The problem did not lie in the reality of our social life, but *epistemological* or *conceptual* construction of the matters involved. Despite Mill’s denial of catallactics, Jevons saw the problem as such.

But we begin with Gossen. His first law of the decreasing utility relates to one commodity only, but the second law of equal marginal utility assumes two and can be seen in terms of catallactics (Gossen 1983: 13–4). Exchange must be made, he wrote, “*in such a manner that the last atom one receives from the other will provide the same amount of value for both*”,<sup>5</sup> thus “*the last atom of each commodity received by every individual will create for him the same pleasure as the last atom of the same commodity received by every other individual*”(100). Although the latter predates the Marginal Revolution by two decades, his excessive elaboration and linear decrement functions made him an underestimated contributor to the crucial innovation in economics. George Stigler once described him as “one of the most tragic figure in the history of economics” or a “profound, original and untrained thinker”(Stigler 1965: 82). Nevertheless, we must remember that he closed the book with an ode for selfishness due to a “bird-eye assessment” of exchange to be analyzed later.

In this beautiful fashion did the Creator know how to remove the obstacle that egoism seems to oppose to the welfare of society and to bring about through this egoism exactly the opposite; *He made egoism the sole and irresistible force by which humanity may progress in the arts and science for both its material and intellectual welfare* (299).

---

<sup>5</sup>Italics in citations are original authors’ unless otherwise noted.

Yet he still accepts saturation and thorough knowledge of the structure and result of exchange, which may not always necessary.

As noted above, William Stanley Jevons (1835–82) had to struggle with the authority of Mill and he exploited catallactics as a crucial weapon.

The point of equilibrium will be... that an infinitely small amount of commodity exchanged in addition,... will bring neither gain nor loss of utility. In other words, if increments of commodities be exchanged at the established ratio, their utilities will be equal.... Thus the ten pounds of corn were of exactly the same utility as one pound of beef, there would be neither harm nor good in further exchange at this ratio (Jevons 1957: 96).

The obtained commodity has greater utility than the one transferred (traded away) because of the scarcity gap, where the quantity of the former increases from left to right and of the latter otherwise (Exhibit 1). If you exchange one of each, you have large subjective utility gain (at quantity  $a$ :  $d - c > 0$ ), and though gain decreases as quantity exchanged increases, you have the greatest gain possible when you exchange  $m$  of them. If you exchanged as much as  $b$ , you would lose something (at  $b$ :  $e - f < 0$ ).

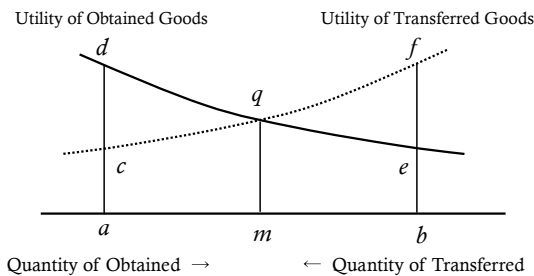


Exhibit 1: Catallactics of Jevons

Although there is a catallactic element in the work of Léon Walras (1834–1910) it is subtle. In *Elements of Pure Economics* (1874), he referred to the stock exchange as a “well-organized market” (Walras 1954: 84), but did so in the context of a barter model. He also wrote that the buyers “make their *demands* by *outbidding* each other” and the sellers “make their *offers* by *underbidding* each other”(83). This is because he assumes an auction, but this is in unusual case in daily experience (either then or now). Although Raphael’s paintings and “a musical snuffbox” with “a magic power” are excluded for their limited supply, securities are “a very special kind of commodity. Furthermore, the use of money in trading has peculiarities of its own, the study of which must be postponed...”(86). On the one hand, he saw competition between Smithian retail sellers like “baker’s, butcher’s, grocer’s, tailor’s, shoemaker’s, etc.” as “poorly organized”. And on the other, “the whole world may be looked upon as a vast general market made up of diverse special markets where social wealth is bought and sold”(84). Thus, while micro exchanges proceed in the less organized markets, economic theory assumes non-existent highly organized one for real goods as for the

macro economy.<sup>6</sup> But we buy bread in exchange for money, not vegetables or meat, and want to buy as cheap as possible while baker wants to sell as dear as possible.

Though he did not introduce marginal principle into utility theory at first (Jaffé 1983: 312–4), he made up a model of many commodities to make the equilibrium “general”, or universal in the scope of goods, to distinguish his model from that of Jevons. Despite his unique contribution in this respect, however, his catallactics differs little from that of Gossen or Jevons.

*We do not have perfect or general market equilibrium unless the price of one of any two commodities in terms of the other is equal to the ratio of the prices of these two commodities in terms of any third commodity (157).*

Although most historians suppose similarity of Carl Menger (1840–1921) to the other two of the triumvirate, his catallactics is *fundamentally different* from theirs. He made two models of market exchange: cattle exchange by isolated farmers and horse exchange for corn in a monopolistic market. In the former, farmer *S* initially has six cows and a horse, whereas *T* has reverse heads of cattle [signs mine]. The utility distribution is shown as the left edge of the table in Exhibit 2 (Menger 2007: 183–6). The farmers give away the abundant cattle to acquire the scarce. Menger declares that exchange takes place only twice and supplements the reason for this judgment in the footnote:

I classify *indifferent* exchanges such as this as definitely *non-economic* since in them the provident activity [Vorsorge, i.e., pre-care] of men is set in motion *aimlessly* quite apart from all the economic sacrifices they may entail (185 [italics mine]).

| 効用表 |    |     |    | 効用表 |    |     |    | 効用表 |    |     |    | 効用表 |    |     |    |
|-----|----|-----|----|-----|----|-----|----|-----|----|-----|----|-----|----|-----|----|
| A 氏 |    | B 氏 |    | A 氏 |    | B 氏 |    | A 氏 |    | B 氏 |    | A 氏 |    | B 氏 |    |
| 牛   | 馬  | 牛   | 馬  | 牛   | 馬  | 牛   | 馬  | 牛   | 馬  | 牛   | 馬  | 牛   | 馬  | 牛   | 馬  |
| 50  | 50 | 50  | 50 | 50  | 50 | 50  | 50 | 50  | 50 | 50  | 50 | 50  | 50 | 50  | 50 |
| 40  | —  | —   | 40 | 40  | 40 | 40  | 40 | 40  | 40 | 40  | 40 | —   | 40 | 40  | —  |
| 30  | —  | —   | 30 | 30  | —  | —   | 30 | 30  | 30 | 30  | 30 | —   | 30 | 30  | —  |
| 20  | —  | —   | 20 | 20  | —  | —   | 20 | 20  | —  | —   | 20 | —   | 20 | 20  | —  |
| 10  | —  | —   | 10 | 10  | —  | —   | 10 | —   | —  | —   | 10 | —   | 10 | 10  | —  |
| 0   | —  | —   | 0  | —   | —  | —   | 10 | —   | —  | —   | —  | —   | 0  | 0   | —  |

Exhibit 2: Catallactics of Menger, A

The *aim* causes exchange. An *indifferent* exchange would be a mere trouble to both. Coming all the way from their rural farm to the city market, they may well demand reward for this physical labor. The context of this model is also vital to understand it: For example, this is placed in the fourth chapter “Theory of Exchange”, where

<sup>6</sup>Walrasian “social wealth” falls in the same category with Millian “phenomena”, while the former’s transaction in the organized market corresponds to Mill’s “operation”. In fact, Walras traced Mill’s criticism of Smith on the definition of economics (Mill 1967: 323), to make people wealthy and to provide the sovereign with revenue (Smith 1976 [1776]: IV.intro.), for it refers not to science itself but to application of it (Walras 1954: 52–3).

Menger begins with criticism of Smith’s “propensity” to exchange and introduces above model in the second section, “Limit of *Economic Exchange*”[italic mine]. This implies that the exchange stops at the second, the last time when exchange remains economic.

The other model assumes farmers buying horses from monopolistic sellers in exchange for grain. The stepwise numbers (Exhibit 3) show that farmer  $B_1$  finds the same utility in the first horse with 80 bushels of grain, and in the second with 70. If three horses are for sale, tells Menger,  $B_1$  purchases two and  $B_2$  one between 60 and 70 bushels, and if six are for sale, then  $B_1$ ,  $B_2$  and  $B_3$  buy three, two and one respectively between 50 and 60 (204–5). Even if a farmer sees the utility level of the  $n$ th horse as same with respective amount of grain this is not an equivalent exchange, since Menger states that every farmer makes exchange for grain whose amount is *between* these numbers.  $B_1$  never buys a horse if the seller demands 80 grain, but buys if he asks 79 or less. If the price is 79, the buyer has a subjective gain by unit grain, and if 75 he gains five. This, therefore, is also an inequivalent exchange. Thus *no exchange will be made at any amount of grain on the table.*

|       | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  |
|-------|----|----|----|----|----|----|----|----|
| $B_1$ | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 |
| $B_2$ | 70 | 60 | 50 | 40 | 30 | 20 | 10 |    |
| $B_3$ | 60 | 50 | 40 | 30 | 20 | 10 |    |    |
| $B_4$ | 50 | 40 | 30 | 20 | 10 |    |    |    |
| $B_5$ | 40 | 30 | 20 | 10 |    |    |    |    |
| $B_6$ | 30 | 20 | 10 |    |    |    |    |    |
| $B_7$ | 20 | 10 |    |    |    |    |    |    |
| $B_8$ | 10 |    |    |    |    |    |    |    |

Exhibit 3: Catallactics of Menger, B

Though these ideas are noted quite clearly in the text, they have attracted little attention of historians<sup>7</sup> and Menger’s position has been described as “cardinal utility theory”. But when we link these to the prices via exchange, the following caution of Menger should be regarded:

However much prices, or in other words, the quantities of goods actually exchanged, may impress themselves on our senses,... they are by no means the most fundamental feature of the economic phenomenon of exchange. This central feature lies rather in the better provision two persons can make for the satisfaction of their needs by means of trade.... To this end they engage in economic activity in general.... Prices are only incidental manifestations of these activities, symptoms of an economic equilibrium between the economies of individuals....

But since prices are the only phenomena of the process that are directly perceptible, since their magnitudes can be measured exactly,... it was easy to commit the error of regarding the magnitude of price as the essential feature of an exchange, and as a result of this mistake, to

<sup>7</sup>An exception is found in an essay by Salerno. Taking up two models of Menger, he wrote: “... [T]he two parties would continue to exchange one horse for one cow as long as the value of the good each received exceeded the value of the good he gave up, that is, as long as the two parties valued the goods they exchanged in inverse order” (Salerno 1999: 94); “For example, at a price of seventy bushels, only  $B_1$ , the buyer, ranks the horse above the purchase price [quantity of corn], while  $A_1$ , the seller, and  $B_2$  and  $B_3$ , the excluded buyers, all rank the purchase price above the horse and are content to depart the market without it”(97).

commit the further error of regarding the quantities of goods in an exchange as *equivalents*. The result was incalculable damage to our science since writers in the field of price theory lost themselves in attempts to solve the problem of discovering the causes of an alleged *equality* between two quantities of goods.... But such an equality of the values of two quantities of goods (an equality in the objective sense) nowhere has any real existence (191–2).

Thus if we see his numbers as the prices of particular goods, we may lose our path. The principle of the agent’s valuation common to both models is that agents deduce a subjective gain from the *inequality* of two goods exchanged. And it is the subjective *utility gap* that assures their selfish aim.

Before providing a formulation for this idea, we establish a comparative catallactics between two camps of the revolutionary four. The difference between Menger and Jevons lies in the pattern of utility decrement and the principle of exchange, whether or not to perform the third. But what matters is the latter. We construct another utility distribution as Exhibit 4 to examine this.<sup>8</sup> Let us define a Net Utility (NU) as “utility gap between the obtained goods and the transferred”. NU is 36 in the first exchange, 18 in the second, and 0 in the third. Nonetheless, a Jevonsian agent will make the third, and suffers from *shortage of selfishness*: The Net Total Utility amounts 36 in the first, 54 in the second, and remains 54 in the third. Though the efficiency inevitably decreases as quantity increases, contribution of the third is null, and it is not efficient. As this conclusion holds regardless of the numbers, we may say that *the problem lies in the very idea of the equivalent exchange*.

| Utility   | 1  | 2  | 3  | 4   | 5   |
|-----------|----|----|----|-----|-----|
| Obtained  | 44 | 32 | 22 | 14  | 8   |
| Alienated | 8  | 14 | 22 | 32  | 44  |
| Net       | 36 | 18 | 0  | Δ18 | Δ36 |

Exhibit 4: Modified Common Utility Measure.

Note: Deficit is represented by “Δ”.

Arguments can be developed against this: (i) There will be little difference between them in larger-unit exchanges, (ii) If divided into very small units, there will be little difference in between, and so infinite division is assumed. Both, however, can be rebutted: (i) Though this thesis is correct in itself, we exchange only one unit in most cases, where the exchange makes no sense since we have no NU or NTU whatsoever, so that we are far from selfish, nor are we rational; (ii) We have no feasibility of division of the goods if they are physically indivisible or lose marketability if divided. This is the case in a great variety of goods: clothes, appliances, books or automobiles, etc. In other words, textbook doctrine, which often supposes foodstuff, is a *special theory of utility*. Even if we assume food, menu cannot be divided as we like. Thus

---

<sup>8</sup>Gossen thought that utility decreases with linear function as quantity increases due to saturation. So he is close to Menger and different from Jevons or Walras, but had influence on the latter. Menger read Gossen after the publication of the *Principles*, and he does not appear there. Jevons did not know the precedence of Gossen in the preface of the first version of *The Theory*, but mentioned in the second version (1879) to him (Jevons 1957: xxxii-vi). Walras acknowledged Gossen in the preface of the first version of the *Elements* (Walras 1954: 36–7).

arguments above can arouse questions in return: Why economic theory presupposes, often without qualification, goods which are divisible and whose division does not affect their value as merchandises?

Further inconvenience may arise. The dominant theory sees at most two aspects, out of three, of the functions of mind and their results: (i) first-person’s subjective valuation, (ii) partner’s similar valuation—this can be omitted if theory assumes that his is symmetrical to the first-person’s, (iii) a bird-eye view on the result of the exchange with respect to the change in utility.

Let us consider these three in the Jevonsian exchange. He refers to the first-person’s valuation only. This causes no problem when the second-person’s utility distribution is symmetrical to him. But if the partner demands four units, no exchange will be made (Exhibit 5). Another cross below the first person will tell clearly the underlying deficiency of the model (Exhibit 6). While the infinite division of goods is seen essential to ensure the utility maximization, that it hinders the realization of exchange must be known first. As the single person view still remains here, Jevonsian model is problematic. While correctly pointing out the difficulty of “double coincidence of wants” in barter (Jevons 1875: 3), he presupposed a far tougher “double coincidence of wants in a specific quantity”, which is self-contradictory.

| Utility   | 1  | 2  | 3  | 4  | 5  | 6  | 7  |
|-----------|----|----|----|----|----|----|----|
| Obtained  | 44 | 32 | 22 | 14 | 8  | 4  | 2  |
| Alienated | 2  | 4  | 8  | 14 | 22 | 32 | 44 |

Exhibit 5: Another Possible Utility Distribution

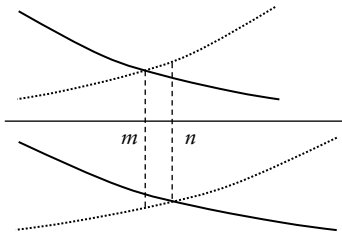


Exhibit 6: Blindness of Jevonsian Catalactics

A Utility Rod will help us understand that this problem can easily be settled in the inequivalent exchange (Exhibit 7). It is a stick like a ruler on whose left-hand side the utility level of agent *S* is marked for two goods (bold line shows that of commodity *A* and thin line *B*) and right-hand side agent *T*’s. The interpersonal comparison of utility, of course, was considered impossible by most twentieth-century economists (e.g. Robbins 1952: 138ff.), but we suppose, to understand the nature of exchange, as if it were possible. The Rod may have some variants: (i) One’s utility distribution crosses that of the other’s, (ii) One’s distribution supersedes the other’s, (iii) One’s distribution is included within the gap of other’s. But exchange can be made in any case, as the requirement for exchange to take place is *simply* that the order of the bold and thin lines differs between two people. This suggests that there are far fewer limitations on Menger’s notion of exchange and it is easier to perform though still limited by the double-coincidence term.



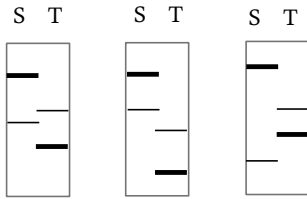


Exhibit 7: Utility Rod

That Menger takes into consideration the valuation of two market participants (Exhibit 2) is crucial. We add to it the bird-eye view. There were seven heads of cattle before the first exchange, when the initial total utility of each farmer was 200 uniformly. The total utility after the second exchange is 260. Micro activity increases macro satisfaction from 400 to 520. Exchange activity increases 120 of Walrasian “social wealth” while the heads of the cattle never increase. The source of this increment can be detected by the bird-eye assessment only: The same old goods were reappraised through the transference from the one who appreciates them less to who does so more.

Gossen was completely aware of this. He showed it in his own geometrical scheme (Exhibit 8). Agent *S* has *d* of commodity *A* and value shown as *abc* of the upper figure, while *T* values *d'* of *B* as *a'b'c'* in the bottom. The utility decreases to reach zero at half the amount, and exchange makes the residual deadstock enjoyed by the partner (Gossen 1983: 95–6 [signs altered]). Thus the same total amount of goods brings two times more utility from the bird-eye perspective.

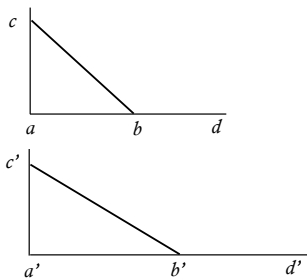


Exhibit 8: Bird-Eye Assessment of Exchange by Gossen.

Source: Gossen 1983: 96.

While any of exchanging “operation” is made from selfish motivation, it serves for the partner *in effect* as “phenomenon”. Then we find that efficient public benefit can be brought about through efficient selfish activity only. Although conventional view tends to support altruistic activity in association with “sympathy” *in ethics*, little attempt has been made to clarify altruism *in exchange*.<sup>9</sup>

<sup>9</sup>There is no such thing as “altruistic exchange” at sacrifice of one’s own benefit. We can imagine an exchange in which *S* provides something for *T*, while *T* provides nothing or far less than *S* does. But the former is a grant and the latter never takes place voluntarily and continuously. If *T* is to be granted, or provided far more than his provision, then a problem should arise which person to be *T*, and strife will be triggered. The one selected should be selfish: otherwise the transaction

## 4 Reorganization of Catallactics by Later Austrians

Ludwig von Mises (1881–1973) made a great progress in utility theory by questioning a series of *assumptions* latent in the traditional argument.<sup>10</sup>

The law of marginal utility and decreasing marginal value is independent of Gossen’s law of the saturation of wants (first law of Gossen). In treating marginal utility we deal neither with sensuous enjoyment nor with saturation and satiety (Mises 1998: 124).

Instead of a *passive* acceptance of saturation, Mises focused on *active* or purposeful deeds. This break with *saturationism* began in the *Theory of Money and Credit* (1912) and opened up an alternative paradigm.

The judgment “Commodity *a* is worth more to me than commodity *b*” no more presupposes a measure of economic value than the judgment “*A* is dearer to me... than *B*” presupposes a measure of friendship (Mises 1980: 58).

We term “saturationism” because while continuous consumption is a matter of choice, economists usually presuppose it without due reason, often referring to the Weber-Fechner’s law of saturation.<sup>11</sup> Even if the law is correct as such, any agent can avoid its actualization. Mises dissociated valuation from it (1998: 125–7). This conversion involved the invention of *another marginal principle*.

The person making the choice does not have to make use of notions about the value of units of the commodity. His process of valuation... is an immediate inference from considerations of the utilities at stake. When a stock is valued as a whole, its marginal utility,... the utility of the last available unit of it, coincides with its total utility, since the total supply is one indivisible quantity (1980: 60).

This “last available unit” can be confused with the textbook saturationist doctrine, which may be dubbed as “the last-unit marginal principle” in continuous consumption, but this is quite different from that. Although mathematical analysis was done

---

itself would be nonsensical. In this case, one’s selfishness may do harm to the other and exchange could not be reciprocal, or mutually altruistic (Note that we do not mean agent’s motivation but *consequence* from objective or third-person’s viewpoint). If so, the simplest solution is that everyone be selfish when it comes to exchange. This does not exclude the possibility of being altruistic in other occasions of our life. And as Mill considered, suffice it to say that economists need not care other aspects of life than the matter of wealth.

<sup>10</sup>Commentators on Mises tend to concentrate on his methodology, especially his “apriorism” with ample doubts (e.g. Gloria-Parermo 1999 and Oakley 1999). But he warns that researchers should concern such concrete but chief economic problems as law of returns, Ricardian comparative cost, economic calculation, etc., without “a most careful examination” of which “nobody can expect to comprehend what praxeology means and what its specific epistemological problems involve” (Mises 2006: xiii). We examine the consumer theory in response to this message. His apriorism is an application of axiomatic precept of the Viennese economics, which is also seen in Morgenstern in a different form (see the last section).

<sup>11</sup>This law has been quoted by representative economists. See, for example, Edgeworth 1881: 59–62; Pareto 2004: IV.33. Especially, Edgeworth introduced the law as the “postulate” and largely assumed physiological reaction of our physical function.

with some reservation in the time of Cournot (1838), Walras took a rigorous view<sup>12</sup> and found market phenomena of the formation of each value “natural” (Walras 1954: 69). Utility theory of Mises stands in stark contrast to this view.

Action does not measure utility or value; it chooses between alternatives. There is no abstract problem of total utility or total value.... [T]here is no such thing as a calculation of [use] values.

An inveterate fallacy asserted that things and services exchanged are of equal value.... The basis of modern economics is the cognition that it is precisely the disparity in the value attached to the objects exchanged that results in their being exchanged. People buy and sell only because they appraise the things given up less than those received. Thus the notion of a measurement of value is vain.... An individual may attach the same value to two things; but then no exchange can result.... Values and valuations are intensive quantities and not extensive quantities. They are not susceptible to mental grasp by the application of cardinal numbers (Mises 1998: 122, 204–5).

We name this paradigm the Pure Qualitative Utility Theory (hereafter PQUT). It calls for the inequivalent exchange principle for any amount of goods are conceived as one to make them serve for agent’s *aim*, and as such this *mentally* bundled single unit of obtained goods must have larger utility than the transferred. If agent *S* exchanges three pears for two apples of agent *T*, observes Mises [signs altered], it is clear that *S* appreciated two apples higher than three pears, knowing nothing of the numerical utility for each fruit (1980: 53), and not being obliged to know it. Thus Mengerian cardinal “counting-up” of each unit’s utility is abandoned. This is natural in so far as no one had ever established cardinal utility of goods or additive theorem thereof (see the next section).

Murray N. Rothbard (1926–95) integrated Mengerian theory and Misesian PQUT.

The mathematical representation of the calculus rests on the assumption of *continuity*.... There is, however, no reason why marginal utility must be conceived in calculus terms. In human action, “marginal” refers not to an infinitely small unit, but to the *relevant* unit. Any unit relevant to a particular action is marginal. For example, if we are dealing... with single eggs, then each egg is the unit; if we are dealing in terms of six eggs, then each six-egg carton is the unit.... To clarify the relationship between marginal utility and what has been misnamed “total utility” but actually refers to a marginal utility of a larger unit, let us hypothetically construct a typical value scale for eggs (Rothbard 2011: 300).

---

<sup>12</sup>Cournot was prudent enough to abstain from disguising accuracy of it: “Since so many moral causes capable of neither enumeration nor measurement affect the law of demand, it is plain that we should no more expect this law to be expressible by an algebraic formula than the law of mortality ...” and “Let us not forget that, strictly speaking, the principle just enunciated admits of exceptions, because a continuous function may have interruptions of continuity in some points of its course”(Cournot 1897: 47, 51). But Walras, who appreciated him very much (Walras 1954: 44), drastically changed the tone: “This whole theory is mathematical.... [T]he proof of the theory must be given mathematically.... It is only with the aid of mathematics that we can understand what is meant by the condition of maximum utility, for... mathematics enables us to see that a given party will obtain the greatest possible total satisfaction of his wants.... Nor can we understand without mathematics why or how current equilibrium prices are arrived at not only in exchange, but also in production capital formation and circulation...”(43).

We call this Misesian alternative the “relevant-unit marginal principle” taking after above term. The Rothbard Chart (Exhibit 9) was drawn from this viewpoint. It tells the existence of a “law of increasing marginal utility” linked to the relevant-unit principle, which has nothing to do with continuous consumption or saturation. Instead, it simply implies that any larger amount of goods, which you may use anytime, has larger utility.

|         |                              |                 |                       |                           |            |
|---------|------------------------------|-----------------|-----------------------|---------------------------|------------|
| 3 eggs  | ↑ Increasing<br>↓ Decreasing | 3 eggs          | <i>Last-Unit Side</i> | 3 cows 120                |            |
| 2 eggs  |                              | 2 eggs          | ▼                     | 2 cows 90                 |            |
| 1 egg   |                              | 1 egg = 1st egg | ▲                     | <i>Relevant-Unit Side</i> | 1 cow 50   |
| 2nd egg |                              | 2nd egg         | ▼                     |                           | 2nd cow 40 |
| 3rd egg |                              | 3rd egg         | ▼                     |                           | 3rd cow 30 |

Exhibit 9: Rothbard Chart, With Its Variants.

Source: Rothbard 2011: 301 [abridged].

Note: Original (left), Decoded (middle) and Re-Integrated (right)

According to this account, there is no such thing as “total utility” since *any necessary quantity is total and marginal at the same time* in the agent’s eye: “marginal” denotes “necessary total for my present aim”.<sup>13</sup> Suppose that your torch goes out and you must get four batteries for it. Do you seek for a commodity A marginal utility of whose last unit coincides with that of the forth battery? Does the second have less utility than the first? Which is the forth at all? How many months does it take to find the thing you are searching for in darkness?

Now it should be clear that textbook catallactics may not be useful for an acting person and that only relevant principle assures selfishness and praxeological rationality, or *rationality of action*, not calculation. The most important factor in the praxeological rationality is *feasibility*. The subjective maximization is useless for it is nothing more than a daydream if the equivalent exchange is not feasible. Anyone who clings to it within the last factor of the Hobbesian Paradigm is least rational. Exchange requires two people at least and they can obtain what they want through Misesian exchange, which implies that two selfish agents benefit each other though both are selfish. This is not a paradox, and there is no split in personality between selfishness and morality now. What matters in economics is not ethical problems but a *mechanism* that emerges out of exchange often beyond each agent’s intention. Selfishness was evil at first. Then it was mitigated and catallactic analysis developed. But as passive acceptance of saturation was assumed, active acquisition that is feasible and its result was overlooked. Again, these essentially new ideas in modern times,<sup>14</sup>

<sup>13</sup>While referring to the utility theory of Mises, Moscati (2013: 21–2) took it for integration of marginal utility to total. But we see marginality as twofold. High and Bloch 1989 and Hoppe 1999 refer to the Rothbard Chart.

<sup>14</sup>What is novel was another marginality. The inequivalent exchange was introduced by Aristotle in *Politics* (1257a) and was mentioned by St Thomas in *Summa* as “pro communi utilitate utriusque [for the *common utility* of both parties]” (1975: 215 [italic mine]). Rothbard took a notice of the latter precedent (2006: 54).

though formulated quite succinctly by Mises and Rothbard, have been discussed very little by historians of economic thought.<sup>15</sup>

## 5 Selfishness and Vicissitude of Utility Theory

Historians of economics generally see a structural change in the way that economists have conceptualized the utility function from early neo-classical cardinalism through ordinalism of Pareto and theories of indifference curve, and on to neo-cardinalism of Morgenstern. The latter turn came about as a result of the interwar debate on the “determinateness of utility function”, which is useful to conceptually reconstruct here. The key to neo-cardinal utility function seems to lie in two equivalent relations which form a logical lattice to represent subjective utility distribution numerically: (i) non-cataclastic equivalence and (ii) cataclastic equivalence.

Vilfredo Pareto (1848–1923) suggested the *equal difference* in utility levels of two food baskets whose amounts vary (Pareto 2014: IV.32). This difference is called “transition”. You have two sorts of goods, e.g. one *A* and two *B*'s. If you acquire another *A* and another *B*, you have two *A*'s and three *B*'s. Let us write the utility level of each basket as  $U_1(1A, 2B)$  and  $U_2(2A, 3B)$ , and transition between them as  $T_1(U_1 \rightarrow U_2)$ , which is identical with  $U(2nd\ A, 3rd\ B)$ . Then suppose that we look for  $T_2(U_2 \rightarrow U_3)$  which assures  $T_1 = T_2$ , or find  $U_3(xA, yB)$  so as to be  $U_2 - U_1 = U_3 - U_2$ . And if  $T_3$  is  $T_3(U_1 \rightarrow U_3)$ , then  $2T_1 = 2T_2 = T_3$ . Because this equivalence is not related to exchange for different sorts of goods, we may call it homogenous or non-cataclastic equivalence.

The other one can be attained via “juxtaposition” or “transcription” between the *equal utility levels of different goods*. When you exchange two apples for three batteries, you juxtapose  $U(2nd\ A)$  with  $U(3rd\ B)$  as far as such a rule as Gossen's second law holds. Battery can be seen as a measure for the utility of apple here. Since this equivalence is found in exchange, it can be called heterogeneous or cataclastic equivalence.

Oskar Morgenstern (1902–77) seems to try to exploit both in the *Theory of Games and Economic Behavior*. Recognizing that the importance of non-cataclastic equivalence was pointed out by Pareto, he sought to find the middle point, as Euclid did in geometry, between utilities of two goods making use of probability-weighted monetary income (Morgenstern and Neumann 2004: 3.3.2). A person shows his preference as regards respectively fifty-percent of combination of incomes *B* and *C*, and non-probabilistic income *A*. If he prefers *A* to *B* and *C* to *A* ( $B < A < C$ ), and *A* to the combination of *B* and *C*, then the transition from *B* to *A* is larger than that from *A* to *C*. If he finds no difference between these two transitions, *A* is the middle point between *B* and *C* (ibid.).

With respect to the cataclastic equivalence, he introduced money as the measure through a “juxtaposed representation”.

<sup>15</sup>Rothbard referred to the utility theory of Mises in another occasion (Rothbard 1999: 149–50).

The conceptual and practical difficulties of the notion of utility, and particularly of the attempts to describe it as a number, are well known.... We wish to concentrate on one problem.... We shall therefore assume that the aim of all participants in the economic system, consumers as well as entrepreneurs, is money.... This is supposed to be unrestrictedly divisible..., freely transferable and *identical... with whatever* “satisfaction” or “utility” [which] is desired by each participant (Morgenstern and Neumann 2004: 2.1.1 [italics mine]).

Thus the neo-cardinal Expected Utility Theory is presented through the lattice-wise logical endorsement of non-catalactic and catalactic equivalences. The idea, first introduced by Fisher (1892: I.4–6), was developed by Franz Alt (1910–2011) in the interwar debate, but he did not use money (Alt 1971: 428–9).<sup>16</sup> His conclusion was indebted to the theory of the continuity of real numbers by a German mathematician Richard Dedekind: Any amount of a given good has equivalence in a particular amount of another good through the catalactic equivalence, and we can find  $U_3$  starting from *any* amount of the former by means of the non-catalactic equivalence.<sup>17</sup> A utility function uniform to positive linear transformation can be assumed in this way. Neo-cardinal, or index, utility schedule has some in common with two co-existing temperature systems: Westerners use Fahrenheit whereas Japanese use Centigrade. But even if the degrees differ, the temperature itself is identical at the same place at the specific time (Exhibit 10). For that reason, Morgenstern thought that this theory could supersede the Paretian indifference curve:

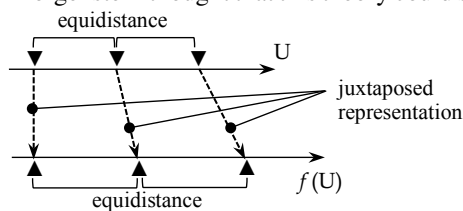


Exhibit 10: Equidistant Utility Transition and Its Transcription.

At any rate we hope we have shown that the treatment by indifference curves implies either too much or too little: if the preference of the individual are not all comparable, then the indifference curves do not exist. If the individual’s preferences are all comparable, then we can even obtain a (uniquely defined) numerical utility which renders the indifference curves superfluous (3.3.4).<sup>18</sup>

<sup>16</sup>Alt was the mathematical tutor of Morgenstern and was recognized by him as an important contributor to the axiomatic theory of utility (Morgenstern 1972: 1165). Morgenstern was very proud of his method: “[O]ur axiomatic system appears to have been one of the first in economics for which an actual, rigorous proof was given”; “It is one of the great pleasure of my life that those few passages on utility theory we wrote have provided so much stimulus for others to concern themselves deeply and in a fresh manner with the notion of utility which is forever basic for any economic theory” (1976: 65, 69). The first chapter of *Theory of Games* was written by Morgenstern (Rellstab 1992: 79, 86).

<sup>17</sup>The point lies in attaining the dyadic relations among given numbers, and this depend on equidistance of transition as Alt emphasized. Thus finding two times larger transition is not different from identifying the middle point.

<sup>18</sup>In fact, Pareto introduced the indifference curve quite intuitively without proof (Pareto 2004: III.52), whereas Morgenstern deduced his utility theory in an axiomatic manner.

Despite this seemingly rigorous attempt, however, we see that the index utility function is *yet to be established*. Morgenstern's demonstration seems to depend on two assumptions: (i) utility of a commodity can be represented by the money price one pays for it in exchange, and (ii) we can find the middle point of utility between two specific sum of money by weighing the probability onto these sums. However, (i) is not attainable. First, fiat money of our time has no utility: No one uses bills as pieces of paper and deposit money has no body. Second, even in the days of commodity money, most people did not have coins to use as metal. And (ii) is not natural taking the role of money in daily transaction into account: As the money we use to buy goods is not exposed to risk, there is little reason for employing the notion of probability here. His theoretical manipulation seems to be deduced to find the middle point to establish the non-catalectic equivalence between two sums.<sup>19</sup>

Remember that Menger denied the juxtaposed representation in goods-money transaction as well as in barter due to the absence of the law of one price in actual market. He wrote:

In my discussion of price theory [cited above], however, I have shown that equivalents of goods in the objective sense of the term cannot be observed anywhere in the economy of men, and that the entire theory that presents money as the "measure of the exchange value" of goods disintegrates into nothingness, since the basis of the theory is a fiction, an error (Menger 2007: 273).

His view is at odds with the EUT.<sup>20</sup> What is more problematic is that arguments of utility seems disappeared from the theoretical front of today's economics, because the EUT transformed the utility from the matter of goods into that of money whose

---

<sup>19</sup>Morgenstern's assumption seems contradictory. If it were the case, a person who bought a coat at half price in a winter sale would have a half satisfaction, and be so sad as to ask the seller price raise to recover his utility foregone. But quite contrary is the case. Some may think that non-catalectic equivalence can be attained by weighing probability onto goods, not money. In fact, he implied this as a digression: "Assume that an individual prefers the consumption of a glass of tea to that of a cup of coffee, and cup of coffee to a glass of milk. If we now want to know whether the last preference—i.e., difference in utility—exceeds the former, it suffices to place him in a situation where he must decide this: Does he prefer a cup of coffee to a glass of content of which will be determined by a 50%–50% chance device as tea or milk" (3.3.2, n.). Nonetheless, this is no more than an idealistic distortion of real experience. He seemed aware of the limit of the EUT: "I know of no axiomatic system worth its name that specifically incorporates a specific pleasure or utility of gambling together with a *general theory of utility*.... [L]et me repeat that in my view the matter of finding a *comprehensive theory of utility is far from settled*.... What the ultimate theory of utility will look like is hard to imagine even after so many works have been published on the basis of the von Neumann-Morgenstern theory" (Morgenstern 1976: 68–69 [italics mine]). This was written in the year before his death. Much praised forerunners may often be less convinced than later followers in what they invented.

<sup>20</sup>While he refers to exchange value, not utility, this does not affect the conclusion. The objection is shared by Jacob Viner, who wrote "If money income be proposed as a measure of the trend of welfare, there is the obvious objection that the monetary unit... is a yardstick which varies arbitrarily in its own length from time to time. If allowance is made, by reference to a perfect index number, for the year-to-year fluctuations in the exchange-value of the monetary unit, there is left virtually 'real' or objective income, the flow of goods and services, as the measure of welfare. This is still inadequate..." (Viner 1925: 643).

utility no one cares. As the theory evolves, the treatment of utility in exchange became less interested in the subjective satisfaction of the agent, and the focus of economics moved away from it, which the term “utility” originally meant. This unfortunate result seems to be attributed to the juxtaposed substitution of money for goods on the assumption that utility of goods can be represented by its money price. Consequently, the PQUT, which deals with utility as such, seems to be one of the most justifiable theories about it in modern economics.

Looking back into those who triggered the Marginal Revolution, however, we find a considerable doubt on the textbook view. Though passed away suddenly by an accident while writing his second economic system, *The Principles of Economics* (1905), Jevons simply stated there:

The final end of the manufacturing and commercial processes is consumption, and it is with utilities, not values, that the consumer has to deal. We must recede to utilities in order to comprehend the matter correctly. The price at which a thing is brought is no necessary indication of its utility (Jevons 1905: 4).

Here, we find an anti-EUT view again. Whether the utility theory made a considerable progress from the 1870s on is an open question. Jevons emphasized utility *inequality* between the goods exchanged.

It will be obvious, after a moment’s thought, that if a thing were as useful to the transferor as to the transferee there could be no advantage in barter or exchange.... Every free act of exchange must imply increase of utility; hence... articles... acquired more utility in the hands of their new possessors (3–4).

His reliance on such French economists as Melon, Condillac, Say, etc. is heavier here than in *The Theory*.<sup>21</sup> But these are whom Menger also drew upon. Thus, British Jevons is far closer to Menger than Austrian Morgenstern is. He emphasizes qualitative utility more than quantitative gain: “Utility is plainly the subject matter of economics from beginning to end. It is *the alpha and omega of the science...*” (6 [italic mine]). He sometimes looks like a purely praxeological economist.

In the simpler acts of barter... the whole thing is a question of direct estimation of utility. When a traveler in savage country gives his pocket-knife to the natives to obtain a supply of food, there is little question of market prices. (ibid.)

Moreover, he introduced taxonomy of goods according to the pattern of use. (1) Durable Consumable Commodities, for those which are durable “so long as they are not put to use”: e.g., salt, soap, tinned meats or fuel. (2) Perishable Consumable Commodities, for those whose durability is lower than the first: fresh meat, fish, etc. (3) Perishable Inconsumable Commodities, for those which perish “of their own accord in a brief interval of time,... not destroyed in a single use”: cut flowers, daily

---

<sup>21</sup>“It is very remarkable that, with few exceptions,” wrote Jevons, “English economists have entirely omitted to notice the groundwork of their own science. This is not, however, the case with the French economists, who, almost from the birth of political economy, have placed a chapter on ‘Besoins [wants]’ at the beginning of their treatises” (1905: 7).



newspaper or railway guide. (4) Durable Inconsumable Commodities, for whose which “we can make use of time after time for many years”: house, cart, carriage, knife, plate, etc. (1905: 22–8).<sup>22</sup>

Yet such a conversion may cast a doubt on the integrity of his position: His account that the Durable Consumable should be used “when their utility is probably at the highest point” (23) lest it deteriorates seems not compatible with that in *The Theory*. In fact, he admits that mathematical maximization theory in this case would be complex (ibid.). At any rate, he never discarded the saturationism: When Americans sell bacon to England in exchange for tea and sugar, they are not “absolutely saturated with bacon” but “less saturated with tea and sugar” (5–6). Thus he stopped short of pondering the matter any further. Diamond, wrote Jevons, can be used as either an ornament or a glass cutter. “With all such distinctions economics has nothing to do. It belongs to other branches of the moral and social sciences to investigate the ultimate effects of actions” (12). It is this that the praxeology concerns. Mises started from where Jevons stopped.

Menger, on the other hand, put so strong a stress on the personal or subjective gain that he used *wirtschaftend* as often as 250 times in the *Principles*. This term is made up of *Wirt*, German “lord” or “husband,” and *-end*, “-ing”.<sup>23</sup> English “economy” is related to *oikonomia*, Greek “household administration”, but lacks *personal* conception. English version of *Principles* uses “economizing” for *wirtschaftend*. Menger’s science is one of *Wirtschaftendeslehre*, or *economizing studies*. While Mises altered the way theory is presented, he directly inherited and even bolstered this essentially Austrian spirit.

Although barter is considered as an activity peculiar to ancient times, recent development of the internet has brought about an increased possibility of barter as it resolves in part the double-coincidence restraint. Thus a question may arise: If you are to perform barter, what type of catallactics would you and your partner choose? Technological progress, that is, can make us face the matter again. Misesian theory seems to be most likely to be chosen *simply because it is the easiest*.

Looking through the arguments on utility and exchange in relation to selfishness, we now grasp the less economic and less selfish nature of the alleged selfish man in textbooks. People take economizing studies for non-economizing studies. This difference is expressed in an unconscious and dim manner in the marginal principle, and this unconsciousness kept researchers from understanding the essence of the underlying problem. Economics that depicts *consciously economizing* agent requires praxeological foundation.<sup>24</sup> Marginality alone does not guarantee subjectivity or selfishness. And this might impede the Marginal Revolution from becoming a Subjective or Selfish Revolution.

<sup>22</sup>This taxonomy is followed by the theory of how to utilize or consume goods efficiently (28–35), which is essentially praxeological and one of the most realistic reflections on the matter.

<sup>23</sup>And *-shaft* is a suffix to make abstract noun.

<sup>24</sup>Misesian catallactics stems from his praxeological axiomatism. If it is so, then modern utility theories may be categorized as follows: i) non-axiomatic measurable cardinalism, ii) non-axiomatic immeasurable ordinalism (Paretian), iii) axiomatic measurable neo-cardinalism (EUT), and iv) axiomatic immeasurable PQUT.

## References

### *Primary Literature*

- Alt, Franz. 1971 [1936]. On the measurability of utility. In *Preferences, utility and demand: A minnesota symposium*, ed. John S. Chipman, et al., 424–31. Orlando, FL: Harcourt Brace Jovanovich.
- Aquinas, St Thomas. 1975. *Summa theologiae*, ed. by Marcus Lefébure, vol. 38: Injustice, 2a2ae, 63–79. London: Blackfriars.
- Aristotle. 1932. *Politics*, ed. by H. Rackham, Cambridge, MA: Harvard University Press.
- Bentham, Jeremy. 1962 [1789]. *Introduction to the principles of morals and legislation*, Works, vol. 1. New York: Russell & Russell.
- Cournot, Augustin. 1897 [1838]. *Researches into the mathematical principles of the theory of wealth*, trans. by Nathaniel T. Bacon, London: Macmillan.
- Edgeworth, Francis Y. 1881. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. London: C. Kegan Paul.
- Fisher, Irving. 1892. *Mathematical investigations in the theory of value and prices*, dissertation, Yale University.
- Godwin, William. 1993 [1793]. *An enquiry concerning political justice*, political and philosophical writings of william godwin, vol. 3, ed. by Mark Philip and Austin Gee, London: William Pickering.
- Gossen, Hermann Heinrich. 1983 [1854]. *The laws of human relations, and the rules of human action derived therefrom*, trans. and ed. by Rudolph C. Blitz. Cambridge, MA: The MIT Press.
- Hobbes, Thomas. 1983 [1642, tr. 1651]. *De cive [Philosophical rudiments concerning government and society]*, ed. by Howard Warrender. Oxford: Clarendon Press.
- Hobbes, Thomas. 1992 [1651]. *Leviathan, or the matter, form, and power of commonwealth, ecclesiastical and civil*, ed. by Sir William Molesworth, Collected Works, vol. 3. London: Routledge/Thoemmes.
- Husserl, Edmund. 1970. *The crisis of european sciences and transcendental phenomenology: An introduction to phenomenological philosophy*, trans. by David Carr. Evanston, IL: Northwestern University Press.
- Hutcheson, Francis. 1971 [1725]. *An inquiry into the original of our ideas of beauty and virtue*. Collected Works, vol. 1, Hildesheim. Germany: Georg Olms.
- Jevons, W. Stanley. 1957 [1871]. *The theory of political economy*, fifth edition, New York: Kelly & Millman.
- Jevons, W. Stanley. 1875. *Money and mechanism of exchange*. London: Henry S. King.
- Jevons, W. Stanley. 1905. *The principles of economics: A fragment of a treatise on the industrial mechanism of society, and other essays*, ed. by Henry Higgs. London: Macmillan.
- Malthus, Thomas R. 1986 [1798]. *An essay on the principle of population*, the first edition (1798) with introduction and bibliography, ed. by E.A. Wringley and David Souden. London: William Pickering.
- Mandeville, Bernard. 1957 [1714]. *The fable of the bees, or private vices, public benefits*, ed. by F. B. Kaye, vol. 1. Oxford: Clarendon Press.
- Mandeville, Bernard, 1957 [1732]. *The fable of the bees, or private vices, public benefits*, ed. by F. B. Kaye, vol. 2. Oxford: Clarendon Press.
- Mandeville, Bernard. 2013 [1732]. *A letter to dion*, with an Introduction by Jacob Viner. Colorado Springs, CO: Create Space.
- Menger, Carl. 2007 [1871]. *Principles of economics*, trans. and ed. by James Dingwall and Bert F. Hoselitz. Auburn, AL: Ludwig von Mises Institute.
- Mill, John Stuart. 1967 [1836]. On the definition of political economy and on the method of philosophical investigation in that science. In *Essays on economics and society*, Collected Works, vol. 4, ed. by J. M. Robson, 309–39. Toronto, Canada: University of Toronto Press.

- Mill, John Stuart. 1965 [1848]. *Principles of political economy, with some of their applications to social philosophy*, Collected Works, vol. 3, ed. by J. M. Robson. Toronto, Canada: University of Toronto Press.
- Mill, John Stuart. 1977 [1859]. *On liberty*. In *Essays on politics and society*, Collected Works, vol. 18, ed. by J. M. Robson, 213–310. Toronto, Canada: University of Toronto Press.
- Mill, John Stuart. 1969 [1861]. *Utilitarianism*. In *Essays on ethics, religion and society*, Collected Works, vol. 10, ed. by J. M. Robson. Toronto, Canada: University of Toronto Press, 203–59.
- Mises, Ludwig von. 1980 [1912]. *Theory of money and credit*, trans. and ed. by H. E. Batson. Indianapolis, IN: Liberty Fund.
- Mises, Ludwig von. 1998 [1949]. *Human action: A treatise of economics*, scholar's edition. Auburn, AL: Ludwig von Mises Institute.
- Mises, Ludwig von. 2006 [1962]. *The ultimate foundation of economic science*. Indianapolis, IN: Liberty Fund.
- Morgenstern, Oskar. 1972. Thirteen critical points in contemporary economic theory. In *Selected economic writings of oskar morgenstern*, ed. Andrew Schotter, 267–93. New York: New York University Press.
- Morgenstern, Oskar. 1976. Some reflections on utility. In *Selected economic writings of oskar morgenstern*, ed. Andrew Schotter, 65–69. New York: New York University Press.
- Morgenstern, Oskar, and John von Neumann. 2004 [1944]. *Theory of games and economic behavior*, sixtieth-anniversary edition. Princeton: Princeton University Press.
- Pareto, Vilfredo. 2014 [1906]. *Manual of political economy: A critical and variorum edition*, ed. by Aldo Nontesano, Alberto Zanni, Luigino Bruni, John S. Chipman, and Michael McLure. Oxford: Oxford University Press.
- Robbins, Lionel. 1952 [1932]. *An essay on the nature and significance of economic science*, second edition. London: Macmillan.
- Rothbard, Murray N. 2011 [1956]. Toward a reconstruction of the utility and welfare economics. In *Economic controversies*, 289–333. Auburn, AL: Ludwig von Mises Institute.
- Shaftesbury, Third Earl of. 1995 [1737]. *Characteristics of men, manners, opinions, times, etc.*, ed. by John M. Robertson, The Shaftesbury Collection, vol. 1. Bristol, UK: Thoemmes.
- Smith, Adam. 1976 [1759]. *Theory of moral sentiments*, ed. by D. D. Raphael and A. L. MacFie. Indianapolis, IN: Liberty Fund.
- Smith, Adam. 1976 [1776]. *An inquiry into the nature and causes of the wealth of nations*, ed. by R. H. Campbell and A. S. Skinner. Indianapolis, IN: Liberty Fund.
- Walras, Léon. 1954 [1874]. *Elements of pure economics, or the theory of social wealth*, trans. and ed. by William Jaffé. London: George Allen and Unwin.
- Whatley, Richard. 1832. *Introductory lectures on political economy*, delivered in Easter term, 1831, second edition. London: B. Felloes.

## Secondary Literature

- Gloria-Palermo, Sandye. 1999. *The evolution of austrian economics: From menger to lachmann*. London: Routledge.
- High, Jack, and Howard Bloch. 1989. On the history of ordinal utility theory: 1900-1932. *History of political economy* 21 (2): 351–65.
- Hoppe, Hans Hermann. 1999. Murray N. Rothbard. In *Great austrian economists*, ed. by Randall G. Holcombe, 223–41. Auburn, AL: Ludwig von Mises Institute.
- Jaffé, William. 1983. *William jaffé's essays on walras*, ed. by Donald A. Walker. Cambridge, UK: Cambridge University Press.
- Lovejoy, Arthur O. 1936. *The great chain of being: A study of the history of an idea*. Cambridge, MA: Harvard University Press.

- Moscati, Ivan. 2013. Were Jevons, Menger, and Walras really cardinalists?: On the notion of measurability in utility theory, psychology, mathematics, and other disciplines, 1870–1910. *History of Political Economy* 45(3).
- Oakley, Allen. 1999. *The Revival of Austrian Economics: A Critical Assessment of Its Subjectivist Origins*. Cheltenham, UK: Edward Elgar.
- Rellstab, Urs. 1992. New insights into the collaboration between John von Neumann and Oskar Morgenstern on the *Theory of Games and Economic Behavior*. In *Toward a History of Game Theory*, ed. by E. Roy Weintraub, 77–93. Durham, NC: Duke University Press.
- Rothbard, Murray N. 1999. Ludwig von Mises. In *15 Great Austrian Economists*, ed. by Randall G. Holcombe, 143–65. Auburn, AL: Ludwig von Mises Institute.
- Rothbard, M.N. 2006. *Economic Thought Before Adam Smith: An Austrian Perspective on the History of Economic Thought*, vol. I. Auburn, AL: Ludwig von Mises Institute.
- Salerno, Joseph T. 1999. Carl Menger. In *15 Great Austrian Economists*, ed. by Randall G. Holcombe, 71–100. Auburn, AL: Ludwig von Mises Institute.
- Stigler, George. 1965. *Essays in the History of Economics*. Chicago: The University of Chicago Press.
- Viner, Jacob. 1925. The utility concept in value theory and its critics: II. The utility concept in welfare economics. *Journal of Political Economy*, 33(6): 638–59.

# Frank Knight on Self-interest



Masanobu Sato

**Abstract** The purposes of this paper were to examine Frank Knight's analysis of the market economy and democratic discussion in terms of the character and function of self-interest, and to clarify the uniqueness of his reflections on self-interest. To this end, the paper organized the significance of the re-examination of Knight's thought as it has developed over recent years in terms of an assessment of the proper role of self-interest in modern society. Knight's position could be described as skepticism towards two types of optimism. The first is the optimism of the next generation of Chicago economists who saw wants as a given and affirmed an economic order based on self-interest. The second is the optimism of the liberal institutionalists regarding the control of self-interest. In contrast, Knight's position was ambivalent, recognizing the need for and difficulty in searching for relatively absolute super-individual norms. Knight's unique position was based on this skepticism. For him, searching for super-individual norms through discussion was necessary, but desperate in the real world. This attitude differed from that of the Chicago economists, who saw wants driven by self-interest as a given, and that of the liberal institutionalists, who were influenced by Dewey's optimistic view that self-interest could be controlled positively. Knight's unique position in the genealogy of self-interest is found in this ambivalent attitude toward both types of optimism, and in his nuanced consideration of a market economy and democratic discussion.

**Keywords** Frank Knight · Chicago economist · Institutionalists · Market economy · Democratic discussion

---

I would like to thank Wade Hands for useful Comment for the early version of this paper. This research was financially supported by the Kansai University Fund for Domestic and Overseas Research Fund, 2018, and by JSPS KAKENHI Grant Number 15H03331.

---

M. Sato (✉)

Department of Economics, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

e-mail: [masanobu@kansai-u.ac.jp](mailto:masanobu@kansai-u.ac.jp)

## 1 Introduction

The purposes of this paper is to examine the twentieth-century American economist Frank Knight's analysis of the market economy and democratic discussion in terms of the character and function of self-interest, and to clarify the uniqueness of his reflections. As shown throughout the paper, Knight's position was different from that of the next generation of Chicago economists, who affirmed a market economy based on self-interest, and from the liberal institutionalists of his time, who were critical of the economic system based on self-interest and optimistic about its control. This paper aims to clarify the implications of Knight's nuanced and ambivalent position.

This paper also traces the recent reappraisal of Knight and again clarifies the implications thereof. Through this work, I attempt to situate Knight's thought in a genealogy of considerations of the proper role of self-interests in modern society.

As some scholars have emphasized (Burgin 2009; Emmett 2009), Knight developed a nuanced account of Western modern liberalism, which consists of the market economy and democracy. In a series of analyses of the market economy in the 1920s, including "The Ethics of Competition" (1923), Knight argues that the maximization of self-interest cannot be the basis for the ethical justification of a market economy; therefore, *laissez-faire* cannot be justified from an absolute ethical perspective. In his argument on democratic discussion, Knight points out the problems with actual political discussion and emphasizes that democratic discussion is more than the mere reconciliation of individual self-interests.

As discussed below, Knight is known as a critic of institutionalists who claimed to control the market economy and is sometimes positioned as one of the founders of the Chicago School of Economics. In recent years, however, aspects of his work have been emphasized as differing from that of the next generation of Chicago economists, and his affinity with contemporary institutionalists has also been noted. Knight's position, which shares similarities with each of these positions but seems fundamentally incompatible with them, has troubled many researchers in the past. While attempting to articulate Knight's nuanced position in the debate over the proper role of self-interests in modern society, this essay also presents a new perspective on recent developments in research on his relationship with his contemporary economists.

Knight's view of self-interest was not the same as that of the next generation of Chicago economists, who normatively affirmed an economic order based on individual self-interest. It also differed from liberal institutionalists, which advocated social control of the self-interested market economy by promoting a shift from personal to social responsibility. Knight seemed to take an ambivalent attitude toward these two positions. As discussed below, Knight's attitude was based on resignation, which means he recognized and understood the complexity of the problem of self-interest in modern society.

Section 2 discusses the implications of referring to recent developments in Knight studies to understand his views on self-interest. With reference to recent Knight studies, the significance of contrasting Knight's attitudes about the character and

function of self-interest with those of the liberal institutionalists of his time and Chicago economists of the next generation is clarified.

Section 3 examines the place of self-interest in Knight's analysis of the ethical foundations of the market economy. The point is that Knight did not consider the individual's wants a given, which is the core of the market economy based on self-interest, but pointed to the need for the growth of wants and exploration of better wants.

Section 4 considers the proper role of self-interest in Knight's view of democracy. The point is that Knight emphasized that democratic discussion is not the same as the adjustment of individual self-interests, and pointed out the need to explore social values beyond self-interest through discussion. This is different from the belief in reconciling and governing self-interests with a given social purpose. However, Knight was also deeply concerned about the fact that political discussion in the real world was far from this ideal.

The final section summarizes Knight's nuanced and ambivalent position on the proper role of self-interests in modern society.

## 2 Trends in the Re-evaluation of Knight and Discussion of Self-interest

In recent years, Knight's works have been re-evaluated not only in the study of the history of economic thought, but also in the study of political philosophy.

### 2.1 Trends in the Re-evaluation of Knight

Knight was one of the leading neoclassical economists in the first half of the twentieth century. He is best known for formulating the theory of perfect competition and analyzing the sources of corporate profits based on the distinction between "risk" and "uncertainty" in *Risk, Uncertainty, and Profit* (1921).

However, Knight's interests were not limited to economics in the narrow sense of the word. From the beginning of his academic career, Knight published a number of articles on philosophy, ethics, and history, and contributed to many major journals in economics and ethics, such as *Ethics*. The subject matter he covered ranged from the methodology of economics and social sciences to economic ethics, and from religious theory to political philosophy.

Positioning Knight in the history of economic thought has long been a difficult issue that has troubled many scholars. Knight was known for his critical attitude toward the New Deal and planned economy, even going so far as contending that the New Deal was "virtual dictatorship" (Knight 1968, 89). He is known to have been critical philosopher John Dewey, who viewed individualistic liberalism as old

and advocated a new liberalism, and of his contemporary institutionalists who were influenced by Dewey to argue for the control of the individualistic market economy (cf. Knight 1936).

As a professor at the University of Chicago, Knight's attitudes had a profound influence on many young scholars who would later become central figures in the Chicago School of Economics, such as Milton Friedman and George Stigler. In Chicago, Knight was much more than an economist; he seemed to be a radical liberalist. Knight as a liberalist in the sense of criticizing the New Deal and planned economy left a strong impression on people in the form of explicit support by Friedman, Stigler, and others, or evasion by Paul Samuelson. This is why he was later described as "the founder of the Chicago School," and to borrow Mencken's expression (Mencken 1971), as a "counter-revolutionary philosopher in economics." Knight also joined the Mont Pelerin Society, an organization founded by Hayek and other thinkers in 1947. These points seemed to position Knight as the founder of the Chicago School which glorifies individual liberty and an individualistic market economy.

However, this old, simplistic assessment of Knight has been revisited and has significantly changed in recent studies of the history of economic thought. In recent years, the multifaceted elements of Knight's thought have been elucidated and the differences between Knight and the next generation of the Chicago School highlighted (Emmett 2009; Medema 2010). There has also been a focus on Knight's institutionalistic claims (Hodgson 2001), and the implications of his ambivalent attitudes about pragmatism have been analyzed (Hands 2004, 2006).

Furthermore, Knight's views on democratic discussion have received much attention. The repeated references to Knight's views in *A Theory of Justice* (1971) by the political philosopher John Rawls have drawn renewed attention, and Rawls's writings in Knight's *Ethics of Competition*, owned by Rawls, have been analyzed in detail (Coker 2015). Political philosophers have evaluated Knight's view on democratic discussion as an original liberal theory (Burgin 2009; Roháč 2012; Lister 2017). Through these reassessments, the similarities and differences between Knight and the subsequent Chicago economists of the next generation and institutionalists of his time have been examined.

## 2.2 *The Relationship with Considerations of Self-interest*

These trends in the reassessment of Knight can serve as a reference point for the unique implications of his analysis of self-interest. As we see later, Knight differs from the Chicago economists who affirm a market economy based on self-interest, because he rejects evaluating the market economy in terms of maximizing want-satisfaction. Knight also criticized the dichotomy of freedom (negative/positive), arguing that the dichotomy of formal/effective freedom is preferable. In this respect too, he differs from the Chicago economists who endorse formal freedom and appears to have more in common with liberal institutionalists.



However, Knight also differs from liberal institutionalism, which criticizes an economy based on individualism and attempts to control the economy socially. Indeed, an aspect of Knight's work could be called institutionalist, in that he argued that individual wants are not a given but a product of the economic system. However, Knight was critical of contemporary institutionalists, who were more optimistic about controlling self-interests, and emphasized the difficulty of doing so and need to discuss about the preferable social purpose.

As we see in the next section and thereafter, Knight's argument is nuanced one. It is inevitable that wants cannot be a factual or normative given, that there needs to be a criterion for value beyond the maximization of one's own interests, and that democratic discussion beyond the mere reconciliation of private interests. However, the above is extremely difficult to achieve in the real world.

### 3 Self-interest as Fact and Criterion

In this section, I refer to the normative analysis Knight developed in the market economy based on self-interest. I will examine Knight's views on individual wants in relation to the defense and critique of the market economy based on self-interest, noting the similarities and differences with the Chicago economists and liberal institutionalists.

#### 3.1 *Knight and the Chicago Economists*

The relationship between Knight and the Chicago School has been rethought in recent years. For example, Medema (2010) points out two contrasting images of Adam Smith in relation to the diversity in the Chicago School, namely are *Chicago Smith* and *Kilkaldy Smith* (Medema 2010). The former resembles the neoclassical economist and can be found in the new Chicago School, such as Stigler and Friedman. The latter is more pluralistic and nuanced, and is what McCloskey labeled the "good old Chicago School," which can be found in Knight, Viner, and Coase (McCloskey 1998). The two are connected to their respective differences based on their assumptions regarding self-interested behavior.

The assumption of self-interested behavior meant something very different in the second half of the twentieth century than it did in the first half, and so is likely to have different content and meaning given to it across these epochs (Medema 2010, 47).

According to Medema, for the new Chicago School of the second half of the twentieth century, markets generally worked well, while governments failed. Therefore, there was no room for government intervention. For the good old Chicago School of the first half of the twentieth century, both markets and the government were imperfect, and the problem was choosing between two imperfect options. The implications of

the latter for government action were less clear (Medema 2010, 47–48). A similar view is expressed by Emmett (2009, 145–155), who presents a composite view of the relationship between the Chicago School and Knight that eschews the simple dichotomy of continuity or disconnection.

As we see later, the idea of not taking the market economy as an absolute, but requiring social standards to properly evaluate it is found in Knight's discussion of the ethical foundations of the market economy. Thus, what are the differences between Knight's view and that of the next generation of Chicago economists? Let us review Knight's stance in terms of self-interest.

### 3.2 *Knight's Ethical Assessment of the Market Economy*

Unlike the old Knight portrait, which associates him with the Chicago School after Friedman, the focus in recent Knight studies has been on Knight as a critic of *laissez-faire*. Knight criticizes the easy justification of liberalism as *laissez-faire* (market economy and political non-interventionism) and questions the ethical foundations of an economic system centered on market competition.

Knight critically examined the ethical justification of the market economic order in a series of essays, particularly in the 1920s.<sup>1</sup> He criticized the efficiency and maximization of wants-satisfaction as a criterion for evaluating the economic order itself. For example, he severely criticized the reduction of the question of ethics to economics (Knight 1922), questioned the existing justification argument of the ethical basis of the market economy (Knight 1923), and developed a critique of economists' utilitarian justification of *laissez-faire* ("maximum freedom") (Knight 1929). His criticism was directed at the efficiency and maximization of want-satisfaction as a criterion for evaluating the economic order itself.

Knight criticized that idea that by viewing the wants of individuals as a given, (1) ethics can be reduced to economics and (2) the market economic order can be "ethically" justified from the perspective of the efficiency of wants-satisfaction. In contrast, Knight pointed out that individuals' wants cannot be regarded as a given because of their nature, human life cannot be considered a mere process of want-satisfaction.

Wants, it is suggested, not only are unstable, changeable in response to all sorts of influences but it is their essential nature to change and grow; it is an inherent inner necessity in them. The chief thing that the common-sense individual actually wants is not satisfactions of the wants he has, but more, and better wants (Knight 1922b, EC: 14).

Therefore, Knight insisted that (1) the problem of an "ethical" evaluation of the economic order can never be reduced to economics, and (2) the ethical evaluation of

---

<sup>1</sup>For example, Knight (1923) not only pointed out that the conditions for perfect competition theory are unrealistic, but also argued that even if the theoretical assumptions of perfect competition were to become a reality, the market economy would not be normatively justifiable.

the economic order must be based on an assessment of the desirability of the nature of the wants it generates, not the efficiency of the satisfaction of wants.

For those to whom ethics is only a more or less “glorified” economics, virtue is correspondingly reduced to an enlarged prudence. However, the essential element in the moral common sense of mankind seems to be the conviction that there is a difference between virtue and prudence, between what one “really wants” to do and what one ‘ought’ to do (Knight 1922b, 29).

This can be understood as an assertion of the need for an independent universe of discourse over the evaluation of the economic order and a value standard distinct from the maximization of individual want satisfaction. From a similar perspective, utilitarianism is also criticized for seeking to maximize individuals wants as a given (Knight 1929).

Here, the difference between Knight and the next generation of Chicago economists comes into play. For the Chicago economists, the individual’s wants as the basis of an economy based on self-interest was a given. The challenge for the market economy was its coordination. However, for Knight, wants were not given in either a descriptive or normative sense. Hence, he did not affirm the market economy for the same reasons as the next generation of Chicago economists.

### 3.3 *Similarities and Differences in Liberal Institutionalism*

Is Knight’s position similar to liberal institutionalism? In fact, as an existing critic of the idea of treating wants as data, Knight named American institutionalists like Thorstein Veblen, Walton Hamilton, and John Maurice Clark (Knight 1922b, 13). In addition, Knight pointed out that the attention to business activities in the manufacture of wants grew because of Veblen’s writings (Knight 1923, 43n). In this regard, Knight and liberal institutionalism appear to have something in common.

Knight agreed with the institutionalists in that the various wants that drive individuals towards economic activity are products of the economic process. However, Knight questioned their tendency to view the expansion of wants and creation of new ones as undesirable (Knight 1922b, 13–14). Knight argued that whether creating new wants is a good thing depends on the nature of the wants created.

Whether it is good or bad to create wants depends on the character of the wants created. One cannot condemn advertising and salesmanship out of hand, unless one is prepared to repudiate most of education and of civilization in general; for most of the desires that distinguish man from the brutes are artificially created. Ethically, the creation of the right wants is more important than want-satisfaction (Knight 1923, 43n–44n).

In other words, it was important for Knight not only to point out that wants are not a given, but also that it makes the question of ethical evaluation inevitable. Thus, Knight argued that wants, which are the basis of self-interested behavior, cannot be a given, either as a fact or criterion, and that human beings seek better wants, rendering the discussion about what is better inevitable.

## 4 The Need to Control Self-interest and the Difficulties Therein

### 4.1 Attention to Knight's Argument About Discussion

In recent re-evaluations of Knight, alongside with the discourse on the ethical foundations of the market economy, much attention has been given to Knight's views on the democratic discussion. Moreover, political philosophers and scholars of economic thought have focused on Knight's argument about discussion.

The impact of Knight's article on the idea of the discussion under the "veil of ignorance" in the "original position," concepts at the core of John Rawls's *A Theory of Justice* (1971), has been analyzed in detail (Coker 2015). A comparative study of Knight's claims and Habermas' discourse ethics has also been conducted (Roháč 2012). Furthermore, others such as Lister (2017) focus on the implications of the criticism of desert theory in Knight's thought.

Of particular interest is the analysis of the significant influence of Knight's writings on Rawls, a leading liberal political philosopher of the twentieth-century (Coker 2015). Rawls, who had been interested in economics since his youth and who in several interviews mentioned that Knight's influence on him, highlighted Knight's argument on democratic discussion as a major influence on the emergence of the idea of the "veil of ignorance" in the original position. According to Pogge, who wrote a biography based on an interview with Rawls in summer 1993, Knight's note on the organization of reasonable communicative states in "Economic Theory and Nationalism" in *The Ethics of Competition* inspired the idea of an "original position" in *A Theory of Justice* (Pogge 2007, 16).

The object of Rawls' *A Theory of Justice* was the tradition of utilitarianism in the West, and at the core of his philosophy was the search for just rules without self-interest in the social contract seen through the theoretical device of the veil of ignorance in the original position. Knight's influence on Rawls is helpful in understanding Knight's stance on self-interest. Rawls is considered one of the leading "liberal" thinkers of the twentieth century, as he was oriented toward redistribution through welfare policies. However, as seen earlier, Knight was often positioned in relation to the opposite: the Chicago School. This again shows that we need to rethink Knight's relationship with the liberal institutionalists of his time.

### 4.2 The Complex Relationship Between Knight and Liberal Institutionalism

Let us re-examine at the complicated relationship between Knight and liberal institutionalism. The American institutionalism of the 1920s and 1930s is known as a methodological reform movement in economics, strongly influenced by Veblen.

However, unlike Veblen, they had a strong orientation toward economic and social reform. The institutionalists were influenced by Dewey's critique of old individualistic liberalism and his insistence on the need to socialize the market economy.

Early liberalism regarded the separate and competing economic action of individuals as the means to social well-being as the end. We must reverse the perspective and see that a socialized economy is the means of free individual development as the end (Dewey 1935, 90).

The concern of institutionalists was the normative problem inherent in the twentieth century, namely the gap between the free pursuit of self-interest and the realization of public benefits. Hence, the agenda was social control of the market economy.<sup>2</sup> In recent years, this aspect of the institutionalist movement has received more attention (Rutherford 2011), and is now considered a contribution to the "new liberalism" of twentieth-century Britain (Furner 2001, 2005).

One of the most famous critics of such liberal institutionalists is Hayek. Although known for his critique of collectivism, Hayek's *The Constitution of Liberty* (1960) criticizes American progressives and liberals. He specifically mentions John Dewey and the American institutionalists. Standing on the negative liberty/positive freedom dichotomy, Hayek strongly criticized Dewey and others for their attempt to view the concept of freedom as a positive one, such as freedom from scarcity, rather than as a negative one, such as freedom from arbitrary coercion. Hayek also criticized institutionalists like John Maurice Clark and Walton Hale Hamilton, who were oriented toward economic control and planning rather than laissez-faire, as "critics of the competitive order" (Hayek 1960, 338).

In *The Constitution of Liberty*, Hayek refers positively to Knight many times as a contrast to the institutionalists; however, Knight's attitude toward liberal institutionalists was not as simple.

Important in this connection is Knight's stance on freedom. Hayek criticized liberal institutionalists in terms of the contrast between negative and positive freedom. In contrast, Knight noted that "such terms as 'formal freedom' and 'effective freedom' would be less question-begging and misleading than 'positive' and 'negative' freedom" (Knight 1943, 180). Knight called attention to the relationship between freedom and the power to do certain things, and said that it is inappropriate to understand freedom simply in the negative sense of a lack of coercion. This appears to be closer to the Dewey-style of liberalism.

But the point here is that, in the long view, the social problem is not so much freedom—liberation—as it is the creation of individuals fit for membership in a free Society (Knight 1943, 181).

However, Knight was extremely critical of the liberal reform claims of his time. To understand this ambivalent attitude, we might consider, as Cowan argues that Knight's argument was strongly on the side of skepticism and criticism of the major

---

<sup>2</sup>Although differences and diversity among institutionalists in the United States have been emphasized in some studies of the history of economics (Morgan & Rutherford 1998), this paper focuses on their similarities and typologizes them.

discourses of his time (Cowan 2016, 183). In other words, Knight was forced to put forth his skepticism of the larger ideas of his time, which may be another problem Knight would make in a different context today, for example.

### 4.3 *Discussion and Self-interest*

“Economic Theory and Nationalism” (Knight 1935b.), which I pointed out earlier as the focus of Rawls’ attention (Pogge 2007, 16), was a relatively long article in *The Ethics of Competition* (1935). This essay attempted to typify political and economic systems for a historical assessment of Western society in the same period.

Knight discussed the historical shifts in three political and economic systems: economic individualism, economic democracy or socialism, and fascist nationalism. He pointed out that liberalism, consisting of a combination of the market economy and democracy, was predicated on the existence of a “frontier”. He pessimistically argued that the market economy in the real world, which has lost its frontiers, creates a cumulative misdistribution of wealth, which undermines the foundations of political democracy based on equal position (one person, one vote), and that in real democratic discussions, persuasion and propaganda, rather than equal discussion, are rampant, and so on.

In light of these historical trends in Western society, Knight pointed out the dangers of the rise of totalitarianism, conformism, and experimentalism in the name of liberalism. For Knight, America’s New Deal was part of this. In critically examining these realities, Knight stressed the importance of discussion. For him, political discussion was not just about mediating individual interests.

The principle of majority rule must be taken ethically as a means of ascertaining a real “general will,” not as a mechanism by which one set of interests in made subservient to another set. Political discussion must be assumed to represent a quest for an objectively ideal or “best” policy, not a contest between interests (1935b, 288n).

According to Coker (2015), Rawls focused on the fourth part of Knight (1935a.) “Social Science and Social Action”. The differences between the social sciences and natural sciences are emphasized here, and it is pointed out that the solving social problems is not a matter of “prophecy and control”. How should the social sciences solve the social problems they are supposed to address? Here again, discussion is important. Knight emphasized that discussion is not simply a process of social members asserting their own interests or of seeking a compromise. The process itself creates/produces an idea that goes beyond self-interest.

The ideal of free society is that social problems should be settled in their large outlines by discussion in which all normal adults participate equally, and in further detail by leadership intelligently chosen by all through public discussion and leading with the intelligent and moral consent and co-operation of the masse (Knight 1935b, 344).

Discussion itself, moreover, is a social activity in the true sense, a “game” which has to be carried on in accord with rules, and these rules must be respected for “absolutistic” or “spiritual” reasons and not from individual utilitarian motives (Knight 1935b, 344)

It is “self-evident” that expression of personal preferences is not discussion and indeed leads definitely toward conflict, that rational agreement involves recognition by all parties of super-individual norms (Knight 1935b, 338n).

This subject was repeatedly discussed in papers from the mid-1930s onward (Knight 1941, 1946). Knight’s emphasis was on the element he found at the heart of liberalism: the continued pursuit of the “relative absolute”. This is the essence of liberalism, which does not regard some authority or standard as absolute, degenerate into a mere relativism that makes the value judgments of various individuals inviolable. If social ends and values are a given, then it is only a matter of finding the means to achieve them, and what matters is the efficiency of the means. However, if social purposes and values are inherently difficult to take as a given, then a social debate over good purposes and values is called for. Knight often described this as the “relatively absolute” nature of value.

For liberalism, no form of value is “absolute”; but this does not mean that one opinion is as good as another, or that error, moral choice, or good taste, is unreal. These things are rather “relatively absolute”; they are valid in the same sense that the objective world is real; i.e., we think in terms of a substance or sub-stratum which may be “eternal and immutable”, but its nature is to be progressively discovered, and never fully or finally known (Knight 1946, 472–473).

This type of discussion does not imply a mere “reconciliation” of individual self-interest. Rather, it is a social activity that follows an absolute, spiritual rule that differ from the utilitarian motives of the individual.

Thus, we can finally understand what Knight saw as problematic about liberal institutionalist claims of the post-New Deal era. For Knight, they interfered with the conditions for open discussion through the exercise of direct control and authority.

Knight often emphasized the difference between discussion and persuasion (Knight 1935b, 1941, 1946). His issue was that in democracy (the political aspect of liberalism) in the real world, “discussion” often becomes “persuasion.” The scientism found in the liberal institutionalism of his time, which can give rise to bad authoritarianism and conformism, appeared to downplay the need for the continuity and openness of critical discussion Knight considered important. Knight thought this could damage human freedom.

In the actual historical transition of liberalism, the realization of Knight’s ideal of discussion was hopeless. This difference between liberalism as an ideal and liberalism in the real world may be the reason underlying Knight’s complex and nuanced attitude.

## 5 Concluding Remarks

The purposes of this paper were to examine Frank Knight’s analysis of the market economy and democratic discussion in terms of the character and function of self-interest, and to clarify the uniqueness of his reflections on self-interest. To this end,

the paper organized the significance of the re-examination of Knight's thought as it has developed over recent years in terms of an assessment of the proper role of self-interest in modern society.

While Knight shared some similarities with both the Chicago economists and liberal institutionalists, he took a different position on the normative evaluation of the market economy based on self-interest and the relationship between democratic discussion and self-interest. Unlike the next generation of Chicago economists, he criticized the ethical advocacy of a market economy based on self-interest and stressed the necessity and inevitability of super-individual norms. However, he disagreed with the optimistic attitude toward the control of a self-interested economy by the liberal institutionalists of his time.

Knight's position could be described as skepticism towards two types of optimism. The first is the optimism of the next generation of Chicago economists who saw wants as a given and affirmed an economic order based on self-interest. The second is the optimism of the liberal institutionalists regarding the control of self-interest. In contrast, Knight's position was ambivalent, recognizing the need for and difficulty in searching for relatively absolute super-individual norms.

Knight's unique position was based on this skepticism. For him, searching for super-individual norms through discussion was necessary, but desperate in the real world. This attitude differed from that of the Chicago economists, who saw wants driven by self-interest as a given, and that of the liberal institutionalists, who were influenced by Dewey's optimistic view that self-interest could be controlled positively. Knight's unique position in the genealogy of self-interest is found in this ambivalent attitude toward both types of optimism, and in his nuanced consideration of a market economy and democratic discussion.

## References

- Burgin, Angus. 2009. The radical conservatism of Frank H. Knight. *Modern Intellectual History* 6 (3): 513–538.
- Coker, David. 2015. Rawls and Knight: Connections and influence in a theory of justice. <http://jepson.richmond.edu/conferences/summer-institute/papers2015/DCokerSIPaper.pdf>.
- Cowan, David. 2016. *Frank H. Knight: Prophet of freedom*. Palgrave Macmillan.
- Dewey, John. 1935. *Liberalism and social action*, G.P. Putnam, Reprinted in Prometheus Books, 1999.
- Emmett, Ross B. 2009. *Frank Knight and the Chicago School in American Economics*. Routledge.
- Furner, M.O. 2001. Policy knowledge: New liberalism. In *International encyclopedia of the social and behavioral sciences*, 299–306. Elsevier Science.
- Furner, M.O. 2005. Structure and virtue in United States political economy. *Journal of the History of Economic Thought* 27 (1): 13–39.
- Gonce, Richard A. 1992. F. H. Knight on capitalism and freedom. *Journal of Economic Issues* 26: 813–844.
- Hands, D.Wade. 2004. Pragmatism, knowledge, and economic science: Deweyan pragmatic philosophy and contemporary economic methodology. In *Dewey, pragmatism, and economic methodology*, ed. E.L. Khalil, 255–303. Routledge.



- Hands, D. Wade. 2006. Frank Knight and pragmatism. *The European Journal of the History of Economic Thought* 13 (4): 571–605.
- Hayek, F.A. 1960 *The constitution of liberty*. Routledge & Kegan Paul.
- Hodgson, Geoffrey M. 2001. Frank Knight as an institutional economist. In *Economics broadly considered: Essays in honor of Warren J. Samuels*, edited by J.E. Biddle, J.B. Davis and S.G. Medema, 64–93. Routledge.
- Knight, Frank H. 1922. Ethics and economic interpretation. *The Quarterly Journal of Economics* 36: 454–481, Reprinted in Knight 1935a, 11–32.
- Knight, Frank H. 1923. The ethics of competition. *The Quarterly Journal of Economics* 37: 579–624. Reprinted in Knight 1935a, 33–67.
- Knight, Frank H. 1929. Freedom as fact and criterion. *The International Journal of Ethics* 39: 129–147. Reprinted in Knight 1947, 3–23.
- Knight, Frank H. 1935a. *The ethics of competition and other essays*, new edition 1997, Transaction Publishers.
- Knight, Frank H. 1935b. Economic theory and nationalism. Reprinted in Knight (1935a), 269–351.
- Knight, Frank H. 1936. Pragmatism and social action. *Ethics* 46: 229–236. Reprinted in Knight 1947, 44–54.
- Knight, Frank. H. 1946. The sickness of liberal society. *Ethics* 56, 79–96. Reprinted in Knight (1947), 440–478.
- Knight, Frank H. 1947. *Freedom and reform: Essays in economics and social philosophy*, 1982 new edition, Liberty Press.
- Knight, Frank H. 1968. The economic principles of the New Deal. In *The thirties: Reconsideration in the light of the American political tradition*, edited by Morton J. Frisch and Martin Diamond, 84–99. Northern Illinois University Press.
- Medema, Steven G. 2010. “Adam Smith and the Chicago School. In *The elgar companion to the shicago school of economics*, edited by Ross B. Emmett, 40–51. Edward Elgar.
- McCloskey, Deirdre N. 1998. The good old Coase theorem and the Good Old Chicago School: A comment on Zerbe and Medemma. In *Coasean economics: Law and economics and the new institutional economics*, ed. Steven G. Medema, 239–248. Kluwer Academic Publishers.
- Mencken, H.L. 1971. Frank H. Knight—Philosopher of the counter-revolution in economics. In *The academic scribblers: American economists in collision*, edited by William Breit and Roger L. Ransom, Holt, Rinehart and Winston. Revised edition, 191–204. Dryden Press.
- Morgan, S. Mary, and Malcolm Rutherford (eds.) 1998. *From interwar pluralism to postwar neoclassicism, annual supplement to Volume 30, history of political economy*. Duke University Press.
- Rutherford, Malcolm. 2011. *The institutionalist movement in American economics, 1918–1947: Science and social control*. Cambridge University Press.
- Roháč, Dalibor. 2012. Knight, Habermas and Rawls on freedom, personhood and constitutional choice. *European Journal of the History of Economic Thought* 19 (1): 23–43.

# Karl Polanyi's Motive of Economy and Institution



Takato Kasai

**Abstract** Even as Karl Polanyi criticized classical economics, he tried to develop a theory that deals with human life. He criticized analyses focused on self-interest exchange; however, he did not deny human selfishness. By presupposing various human selfish acts other than the acquisition of commodities, he identified elements of economies without a market such as reciprocity, redistribution, and householding. In addition, non-exchange economies existed not for altruism but for suppression of individual selfishness by the institutional environment. Polanyi considered the self-interest of the liberal states and international finance as crucial for institutions in the nineteenth century, in which selfish economic behavior functions as a self-regulating market. By supposing the selfishness of nations and individuals, he pictured a world where the self-regulating market defined human economic activities, politics, and social systems. It was his agenda to inquire into a social institution that can restrict individual and national self-interests on a global scale for those who desired international peace through a harmony between economy and society.

**Keywords** Individual behavior · Nineteenth-century civilization · Non-exchange market · Social existence · Subsistence

## 1 Introduction

This chapter discusses the role and self-interests of institutions in the work of Karl Polanyi (1886–1964), an economic anthropologist in the twentieth century. Polanyi was a critic of classical economics. His work resonates even today (especially, after the year 2000, when people could recognize the limits of economics based on neoliberalism) because by expanding economic liberalism through a historical discussion, he identified a destabilized society. Additionally, people believe his assertions can supply a theoretical foundation against the present-day system.

---

T. Kasai (✉)  
Doshisha University, Kyoto, Japan  
e-mail: [t.kasai@outlook.com](mailto:t.kasai@outlook.com)

It is important to consider the self-interest of social institutions to understand Polanyi. His economic argument describing the social reality of humanity was at variance with that of Adam Smith who considered individual self-interest as human nature. Polanyi questioned the market, which decides commodity price and quantity, and inquired into the appropriate relationship between society, humanity, and economy in *The Great Transformation* (hereafter *TGT*). *TGT* was written against the backdrop of the political bloodiness of the World Wars and the economic uncertainty of the Great Depression, which caused mass unemployment. According to Polanyi, these historical dangers emerged from the prosperity of a market whose economics were founded on individual self-interest. Moreover, market limits usually appear despite the claims of mainstream theory. Based on this recognition that an economy can induce historical events, comprehending self-interests as a foundation of people's behavior in a market will shed light on Polanyi's perspective of the world.

We will consider whether humans naturally pursue their self-interest based on Polanyi's discussion. His evaluation of self-interest was viewed as negative because he criticized a market society in which the market decides the commodity price and quantity. Certainly, he criticized economics focused on self-interest exchange. As self-interest exchange became available in the market, Polanyi's negative evaluation of self-interest emerged from comparing it with the altruism of non-exchange economies (Pearson 1977). He hesitated at the idea of people who act according to the rule of individual utility maximization instead of social solidarity in society (Baum 1996). However, Polanyi's view of humanity having no self-interest differs from what he disliked about individual self-interest as a foundation of economic behavior. He was interested in how to manage contemporary society by interpreting past knowledge (Tandy and Neal 1994). As a part of society, he articulated humanity as a way of strengthening society to challenge selfishness as the foundation of economics (Rogan 2017). Moreover, factors such as religion and the environment could affect human behavior according to Polanyi. This chapter will examine the relationship between individual action and institutional behavior by focusing on nineteenth-century civilization created by the market society.

## 2 Polanyi's Theory

Polanyi distrusted situations where the economy usually dominated society and explained the necessity of designing a free society that ensured humanity's existence by prioritizing society over the economy. Specifically, he examined the economy embedded in society, rather than society embedded in the economy. The idiosyncrasy of capitalism, as it was called in nineteenth-century civilization, was revealed by three integrated economic forms that differ in exchange. Further, he illustrated humanity's social existence<sup>1</sup> in the community (Stanfield 1986).

---

<sup>1</sup>Polanyi was enticed by the idea of humanity as a social existence by reading *Hamlet* at various times when he participated in WWI (Polanyi 1954). After returning from the war, he fell sick

Polanyi's three integrated economic forms are reciprocity, redistribution, and householding. These were discovered after considering the works of Malinowski, Thurnwald, and Aristotle, who observed the economy as experienced by humanity. The integrated economic forms can classify various human activities from an economic perspective. He also criticized economic theories that focused only on an exchange as an object of analysis and the market as the sole location for the economic activity. He was especially critical of Malthus and Ricardo<sup>2</sup> due to their contribution to the development of economic theory, which was skewed toward an analysis that did not incorporate humanity. Polanyi indicated that not only exchange but also reciprocity, redistribution, and householding comprise the human economy.<sup>3</sup> These economic forms can exist in any time, culture, or society; the economy, characterized by human behavior, cannot be articulated only as market exchanges.

The reciprocity concept—the first integrated economic form—involves a give-and-take of usable materials with specific partners or groups. This signifies neither simultaneous transactions nor equality in usable materials but rather a mutually beneficial relationship. According to Polanyi, various action principles can result in effective economic activity through institutional patterns; for example, the pattern found in reciprocity is its symmetry. Reciprocity is informally organized and is, sometimes, characterized by ritual and obligatory relationships. In addition, Reciprocity operates in blood-tied organizations, such as family businesses, because their social relationships decide their partnerships.

Reciprocity is ethnocentric when one group acquires usable material from other groups. However, it indicates activities encouraged by an institution centered on people and differs from primitive exchanges such as barter. In Kula trade, a popular example of reciprocity, such usable materials like armlets and chokers circulate among constituent island members in a specific direction. The giver differs from the taker, and the time of giving also differs from the time of taking. Additionally, this symmetry does not solely function between multiple parties without records. Kula trade excludes stock and the accumulation of value since the gifts move in a single circular direction. Goods, commodities, and materials relocate as society decides what people should do with them. Reciprocity implies that a system generally functions by each movement of goods or materials depending on the circumstances surrounding people.

---

and nearly died from melancholy as he considered human existence in a community (Brie 2017; Polanyi-Levitt 2018; Stanfield 1986).

<sup>2</sup>According to Polanyi, Malthus and Ricardo omitted moral aspects that differ from Smith during their development of the political economy; thus, Polanyi again attempted to approximate his economic theory to Smith's insight (Rogan 2017). Polanyi, however, never regarded Smith highly. Instead, he was against the censures in relation to the propensity of exchange and the Act of Settlement and Removal as a basis of the labor market.

<sup>3</sup>Polanyi focused on just the theory of Malthus and Ricardo when he critiqued economics. Reference to Mill and Marshall are rare. Moreover, Keynes's insistence was used only for indicating an example of a big economy which was not borne by free individuals such as the building of the Egyptian Pyramid in ancient society.

Redistribution, by contrast, involves the division of usable materials into constituent members after first collecting them at specific locations or from specific people. This economy, as exemplified by the formation of hunts and festivals in archaic communities, has centrality as its institutional pattern. It also operates by obligatory transfers. Redistribution allocates the materials accrued by a group to individuals by relying on an authoritative power system, such as a chief or a king. The members can acquire materials by depending not only on their contribution to hunting or gathering but also based on necessity while obeying communal custom. Therefore, individuals could avoid starvation by belonging to the community. In addition to a relationship with a neighbor such as a central, political, or religious authority, redistribution locally affects and requires the formation of a community in advance before an economic activity. It is not necessary for all people to have the same social status. One peculiarity of reciprocity is that a person of influence is a mediating agent. Redistribution invites reallocation of materials by using materials stored during war or famine; it usually enhances social ties. Simultaneously, the political power of authorities aggrandizes, forms, and composes social institutions. Certainly, the store of materials does not indicate wealth accumulation for economic growth. As redistribution prominently entered modern nations in the form of a tax system, the social function of redistribution was intermingled with an economic one.

The third principle—householding—indicates a self-sufficient mode of consumption and features a closed group that produces and stores products to satisfy its members' wishes regarding a way of life. It emerges from the relationship between persons comprising a natural unit of humans living together. An autarky as an institutional pattern describes the self-sufficient nature of householding. Polanyi believes householding developed more recently, compared with reciprocity and redistribution, as the economy of mankind. Moreover, he supposed that householding appeared after sufficient agricultural development. He posits that production for self-use does not assume an individualistic and primitive humanity. Additionally, such a supposition is simply a product of modern times. He derived householding as a principle of self-sufficiency from Aristotle:

Aristotle insists on production for use against production for gain as the essence of householding proper; yet accessory production for the market need not, he argues, destroy the self-sufficiency of the household as long as a cash crop would also otherwise be raised on the farm for sustenance, as cattle or grain; the sale of the surpluses need not destroy the basis of householding (Polanyi 1944: 56).

Householding consistently includes a small economy without the propensity to expand, and this differs from money-making. Thus, if an economy uses currency for payments,<sup>4</sup> the economy can be considered as householding if no expansion occurs. Polanyi notes that whether money is used is never a core problem.

Exchange, as the fourth principle, involves a market economy that decides the price and quantity of commodities whereby people trade and transport usable materials.

---

<sup>4</sup>The term of payment indicates the usual diction of economic mediation, although Polanyi discriminates the means of exchange from payment as an obligation (Polanyi 1977).

According to Polanyi, exchange coincides with trade, and transportation is an action for which the economic principle is underwritten by market patterns. Characteristically, market patterns can create a specific institution beyond the role of economic activities. On the one hand, the symmetry, centricity, and self-sufficiency in closed groups form nothing. On the other hand, the “market pattern . . . being related to a peculiar motive of its own, the motive of truck or barter, is capable of creating a specific institution, namely, the market” (Polanyi 1944: 60). The market economy's system dominance and the substantial effects on the entire social organization imply a society embedded in an economy.

Polanyi did not criticize the principle of exchange in economic activity but perceived exchanges based on economic self-interest for profit as problematic. Exchange can exist even in archaic societies. Thus, the issue is not the existence of exchange but a balance among the four institutional forms. In the context of human history, reciprocity and redistribution existed in archaic societies as well as empires. Certainly, humans acted in a householding economy. Exchange, trade, and transport as individual actions for profit could also exist in an archaic society, although they may not have been fundamental to the era. According to Polanyi, the silent barter and armed trade systems did not develop into a market relative to the origins of the exchange. Without a propensity to expand, the local market was created from peace in the archaic society. Further, Polanyi indicated the necessity of previous consolidations of social relationships to exchange goods and commodities in the market and invalidated the assertion that the exchange among free individuals had formed an economy. The idea of free individuals is fictitious in real economic activities. Substantive economy among people in a social existence is tangled with their relationship with their neighbors.

For Polanyi, the market society is seemingly unusual in an embedded society in that the market's logic prevails against society. Although humanity had not experienced a market society until the eighteenth century, the market's foundation—in which people are isolated as individuals—has progressed through devastated communities. This is a problem presented in humanity's economic domination. Therefore, Polanyi's discussion identifies humanity's isolation as a social existence to be caused by economic liberalism (Backert 2009; Dale 2010; Rotstein 1990).

### 3 Self-interest in the Market Society

This section discusses how Polanyi addresses self-interest. His negative evaluation of self-interest has been considered obvious, as he emphasized reciprocity, redistribution and householding in addition to criticizing economics focused on exchange based on the self-interest of individuals. Thus, to overcome individualism in the market society of the nineteenth century, Polanyi affirms Christian socialism because of the recognition of societal man. The economy's embeddedness in society equals its embeddedness of economic self-interests:

The true criticism of market society is not that it was based on economics—in a sense, every society must be based on it—but that its economy was based on self-interest. Such an organization of economic life is entirely unnatural, in the strictly empirical sense of *exception*. ...[T]he self-regulating market of the nineteenth century turns out on closer inspection to be radically different from even its immediate predecessor in that it relied for its regulation on economic self-interest. *The congenital weakness of nineteenth-century society was not that it was industrial, but that it was a market society* (Polanyi 1944: 257–258; italics added by Polanyi for emphasis).

Self-interest is a market society's weakness. However, it is insufficient to merely consider an exchange originating from self-interest in the market as grounds for an argument of Polanyi's negative attitude toward self-interest. Certainly, Polanyi focused on a market society in which an economy based on self-interest became a core feature, but he did not deny humanity's self-interest. The Speenhamland system, which covered deficient wage and ensured a right of life regarding labor, embedded economy in society. The collapse of the system created a labor market characterized by wage, which led to a market society. Self-interest in labor became a protective social movement in an attempt to repress the market's expansion.

The abolishment of Speenhamland was the true birthday of the modern working class, whose immediate self-interest destined them to become the protectors of society against the intrinsic dangers of a machine civilization (Polanyi 1944: 105).

Self-interest of people, on the one hand, is a foundation of market exchange. On the other hand, when people face a survival threat, it can become a basis of movements that can suppress the market mechanism. Self-interest of humanity will become both an accelerator and a brake. Polanyi perceived Smith's self-interest as implying nothing more than a principle of action, and not as a foundation for moral and political duty. Polanyi believed that Smith treated wealth as subordinate in society in addition to treating it as an independent and separate field of study:

In his [Smith's] view nothing indicates the presence of an economic sphere in society that might become the source of moral law and political obligation. Self-interest merely prompts us to do what, intrinsically, will also benefit others, as the butcher's self-interest will ultimately supply us with a dinner. A broad optimism pervades Smith's thinking since the laws governing the economic part of the universe are consonant with man's destiny as are those that govern the rest. No hidden hand tries to impose upon us the rites of cannibalism in the name of self-interest. The dignity of man is that of a moral being, who is, as such, a member of the civic order of family, state and 'the great Society of mankind' (Polanyi 1944:117).

Polanyi calls the analysis of exchange in a market as formal economics. By comparison, he calls the analysis of an entire economy that depicts real human life as substantive economics. Formal economics analyzes the exchange formed by self-interest, which supplied the concepts of morality and duty in the market society of the nineteenth century. Polanyi was interested in wealth, which departed from its position as a part of communal life and acted as a false origin of morality and ethics in the market society. Although the exchange in formal economics connects self-interests with a market society, such a strong connection was rare in the history of humanity

from Polanyi's point of view. Thus, corresponding to his view, these three elements (exchange, self-interest, and market society) should each be discretely analyzed. By considering self-interest separate from the integrated forms, the diversity of such selfish actions will emerge without consolidating it in the exchange.

Polanyi did not locate self-interest by facing it off against three integrated forms. Humans with a natural affinity to altruism do not appear in his work. Reciprocity and redistribution without gain in an archaic society, at a glance, seems to be economic activities operated by altruism as compared to the market society. This includes motives for gain by rational economic individuals. However, if we assume that humanity acted altruistically in archaic societies, why could people not act only through self-interest in the market society? If humanity has become selfish, the reason this occurred should be discussed from Polanyi's perspective. Human nature does not change across times in his discussion. Consequently, he notes that where self-interest is the sole origin of human activity, a negative change in the human spirit, like a devolution, did not appear. Rather, Polanyi stressed on "the similarity of men at all stages," that is, "how humanity has remained the same throughout the ages" (Polanyi 1944: 278). Clearly, Polanyi did not consider the archaic society or the archaic human as ideals.

Polanyi's critiques focused not only on the transfiguration of human nature, but also the social surroundings obeying the development of economic theory, and those who accept such theory as correct. He recognized self-interest as a premise of both reciprocity and distribution; specifically, although each person is selfish, they can recognize further social bonds beyond this.

Reciprocity shifts the emphasis from the utilitarian element of selfish advantage to the warmth of experience and gratification that flow from mutually honorific neighborhood contracts with those to whom we are joined in specific relations of objective status and personal intimacy. Redistribution strengthens internal communal ties by all the psychological means at men's disposal. In addition, he did simply not deny self-interest built-in naturally to humans as it is. Considering with his reference to Smith's self-interest and human generality, the exchange and householding also existed in archaic society because humanity has self-interest at any time (Polanyi 1977: 60).

Regrettably, Polanyi misleadingly explained the non-exchange economy, in which reciprocity and redistribution conform to human nature, with a strong emphasis on critiquing the exchange economy. However, as the four integrated forms of economy subsist in any society, the core issue is their balance. The economy becomes embedded in society due to a loss of balance in the market society.

Additionally, the non-exchange economy does not come into existence through altruism or unselfishness. Any person has a utilitarian side not only in a market society but also in an archaic society.<sup>5</sup> Polanyi posited that "reciprocity demands adequacy of response" without denying Malinowski's argument, implying that no gifts are gratuitous (Polanyi 1957a: 73). People act reciprocally to gain honor and fame and meet obligations in a community. People may sometimes enter slavery if they did not fulfill

---

<sup>5</sup>While Polanyi accepted self-interest as a part of human nature, he denied Bentham's insistence which articulated the social system by stressing on self-interest as the core of humanity.



their reciprocal obligations in society. Polanyi challenged an agenda regarding how well humanity moves in real societies where selfish people typically live. Reciprocity and redistribution then become real not to bless their economic counterparts, but to affirm their actions in finely tuned situations. This signifies an embedded economy based not on altruism but rather social usefulness (such as peace, prosperity and the strength of ties in a community) and selfishness. Polanyi called these beneficial and gratuitous gifts, and he does not consider a creed of the unselfish archaic man who acts for other people. Volunteering and free-cost differ from self-sacrifice; that is, human as social existence. Polanyi does not profusely glorify the past. People can change and are affected depending upon their surroundings: “man can be as good or evil, as social or asocial, jealous or generous, in respect to one set of values as in respect to another” (Polanyi 1944: 49).

Polanyi examined the essence of relationships among men and their social surroundings by focusing on social usefulness as well as institutions without economic efficiency. Polanyi’s ideas regarding humanity have a constancy of accommodation since people can adjust themselves to various circumstances. Additionally, Smith’s propensity of exchange is not necessarily required for Polanyi’s ideas regarding humankind. The balance of forms of economic activities vary in any society because of peoples’ diversity. However, Polanyi demonstrated humankind’s discordance with a market society based on economic theory, which has been revealed as a study on exchanges.

#### 4 Selfish Motives and the Economy

This section will discuss the selfish motives of economic activity, as characterized by self-interest. Polanyi presented three selfish motives in an economic action: subsistence (preservation of life), status (preservation of social relationships among people), and gain (acquisition of money).<sup>6</sup> The motive of subsistence induces a man to act to ensure his survival. Fear of starvation invokes the motive of subsistence, which primitive humans tackled to overcome. The motive of subsistence leads to hunting and gathering to overcome the fear of starvation in the individual acquisition of useful materials. In some situation, working in groups might make things easy. Thus, man generates a social existence by creating groups with other people, and a community plays into this economic behavior as an assemblage to strengthen these social ties.

The motive of status induces humans’ social existence to partake in economic activities aimed at sustaining and improving an individual’s position in a community. The social status of those involved in transactions is connected with economic activity, which is manifested as both an obligation and a public service in which people attempt to change their status to a more profitable position by gaining admiration and honor. Naturally, actions through the status motive primarily appear as

---

<sup>6</sup>Polanyi used the phrases of profit motive and economic motive to mean the same thing as the motive of gain. The motive of gain was mainly used in *TGT*.

reciprocity and redistribution without the reduction of material profits. Polanyi noted that “[i]f the ‘status motive’ is reinforced by that of material benefit, the latter does not, as a rule, take the form of gain made on exchange” (Polanyi 1977: 84). Additionally, Polanyi recognized the status motive as an obligation of a person of distinction; specifically, “[t]he higher-status groups are usually expected to act from honorific motives of duty, obligation and self-respect, but the lower-status groups may well be encouraged to indulge in gainful occupations which are looked down upon, and hardly keep their man” (Polanyi 1977: 84). Status leads to economic activity and, at the same time, defines what a man should do.

In contrast, the motive of gain becomes the most important factor in actions to obtain economic profit in the form of exchange. This motive dominates a smaller part of humans’ self-interest, and “gain and profit made on exchange never before played an important part in human economy” (Polanyi 1944: 45). However, the motive of gain unnaturally comes to the forefront as an action principle in a market society. It was brought about by a historical transformation from an agricultural to an industrial society:

The transformation implies a change in the motive of action on the part of the members of society; for the motive of subsistence, that of gain must be substituted. All transactions are turned into money transactions, and these, in turn, require that a medium of exchange be introduced into every articulation of industrial life (Polanyi 1944: 43–44).

The introduction of currency in an economy does not directly connect to the motive of gain as indicated by the stone money used in archaic societies. The importance of the gain motive expanded by emphasizing monetary transactions in the transfiguration to an industrial society. Previously, humankind typically lacked an idea of gain and wealth:

[W]e only intended to show the manner in which so-called economic motives spring from the context of social life. For it is on this one negative point that modern ethnographers agree: the absence of the motive of gain; the absence of the principle of labouring for remuneration; the absence of the principle of least effort; and, especially, the absence of any separate and distinct institution based on economic motives (Polanyi 1944:49).

Thus, a non-exchange economy appears from the self-motivation for status and subsistence without the motive of gain. Polanyi evaluates self-interest as neither positive nor negative, but as originating from economic activity. This becomes problematic in how an economy becomes congested by the exchanges in a society composed of selfish individuals.

The subsistence motive, such as the fear of starvation, weakens by repeating social relationships in a non-exchange economy created by selfish motives. The subsistence motive can create feedback to depress the individual when it is satisfied. According to Polanyi, this occurs in a process in which man creates a community and breaks free from an individualistic fear of starvation, as exemplified by the non-exchange economy:

The individual’s economic interest is rarely paramount, for the community keeps all its members from starving unless it is itself born down by catastrophe, in which case interests

are again threatened collectively, not individually. The maintenance of social ties, on the other hand, is crucial. First, because by disregarding the accepted code of honor, or generosity, the individual cuts himself off from the community and becomes an outcast; second, because in the long run all social obligations are reciprocal, and their fulfillment serves also the individual's give-and-take interests best. Such a situation must exert a continuous pressure on the individual to eliminate economic self-interest from his consciousness to the point of making him unable, in many cases (but by no means in all), even to comprehend the implication of his own actions in terms of such an interest. This attitude is reinforced by the common catch or sharing in the results of some far-flung and dangerous tribal expedition (Polanyi 1944: 48–49).

Human activity based on the subsistence motive includes the prevention of famine in a community, which ensures human life within social relationships. A community liberates people from the fear of starvation and converts individual problems to collective ones. Preservation and strengthening of social bonds enable humankind's easy survival, just as exclusion from a community leads to the individual facing a fear of starvation. Adversely, belonging to a community preserves humanity, except for a substantial disaster eliminating the community. Additionally, the status motive can appear to preserve entitlement for a community's members, which occurs through the unidimensional elimination of economic self-interests by promoting reciprocity and redistribution. Individual self-interests become embedded in a community. Weakened self-interests are sometimes perceived as altruistic actions. However, these merely express the reciprocity and redistribution for one's interests as subsistence.

Similarly, Polanyi perceived altruistic motives in economic activity as not the core of humanity. His explanation indicates that self-interest involves reciprocity and redistribution in the generation and development of a non-exchange economy. Economic actions stemming from a selfish nature are seemingly altruistic at first glance due to the non-exchange economy through social relationships in a community.

A false dichotomy comes to light: On the one hand, the motive of gain creates an exchange, but on the other hand, the motive of subsistence or status creates reciprocity and redistribution. Additionally, not all exchanges solely occur from the motive of gain: some occur through the status motive and a non-exchange economy borne by the motive of gain. Certainly, these do not comprise a significant part of the economy.

In the case of an exchange, the economy exists relative to the motive of status "to draw a relationship closer and strengthen the ties between the partners," except for the motive of gain (Polanyi 1977: 54). This economy aims to neither eliminate the fear of starvation nor acquire goods for a higher profit. However, man indirectly avoids such fears by strengthening communal bonds. Such social bonds are augmented to overcome individual fear through the exchange economy, which provides a premise for the community's existence. Exchange, originally, is a human economic behavior which does not need the motive of gain:

[E]xchange, the most precarious of human ties, spreads into the economy when it could be made to serve the validation of the community. In effect, economic transactions become possible when they could be made gainless (Polanyi 1977: 61).

Polanyi also referred to reciprocity that uses social relationships to enhance one's profit and gain. This reciprocity purely aims for profits, and differs from the subsistence perspective, which involves the enhancement of social bonds for communal interests. For instance, merchants in archaic societies never chose to acquire profits from exchange economies under powerful authorities: kings and temples. Status defined who does what. From this perspective, rather than such an exchange, reciprocity under a power system considers adventure as valuable and expands traders' profits (Polanyi 1977: 84).

Reciprocity can coexist with the motive of gain in Polanyi's view. Specifically, as "he who trades for the sake of duty and honor grows rich, while he who trades for filthy lucre remains poor" (Polanyi 1957b: 259)—motive of gain did not appear at the forefront. Although an economy for status is irrational at a glance, with a focus on the economy for duty and honor, we could consider it as a profit-oriented economy surrounded by a social instruction and an environment that portray the economy as close to irrational. Therefore, the archaic man acts rationally. Accordingly, the exchange and reciprocity appear from selfish motives and sometimes occur from motives of both gain and status. They differ in whether people act for higher profit or duty and responsibility in encouraging their transactions. In an archaic society, behaviors aiming for obligation or duty could have led to more wealth than a gain-oriented action.

Then the motives of gain, subsistence, and status—which provide the foundation of economic activity—are all selfish. At the beginning of human existence, the individual archaic man connected and cooperated with others to overcome a personal fear of starvation and, specifically, created groups or communities. By building community-based social relationships, the fear of starvation transformed from a personal agenda to that of the congregate. This congregate agenda and group living satisfy the motive of subsistence; humankind's social existence already emerges here. Humankind then enforces these social relationships to secure survival. Alternatively, the selfish motive of status usually involves both reciprocity and redistribution to improve an estate in a community since people's status can become heightened or decline to slavery, which is contingent upon methods in the non-exchange economy. Thus, reciprocity and redistribution appear from selfish motives. The non-exchange economy can initially be perceived as unselfish, regardless of how the actual processing occurs. When the motive for economic activity does not directly correspond to the integrated forms of economy, this leads to a more complicated understanding than in previous suppositions.

Moreover, variety in economic transactions was formed during the Middle Ages in Polanyi's work. The transformation to a market society from this economic variety occurred at the end of the eighteenth century when the labor market appeared. The Speenhamland system's collapse converted humankind from a social existence to individuals who changed under a shared fear of starvation as a social interest. Specifically, people were motivated by subsistence as they again individually faced the fear of starvation in a market society.

Polanyi insisted it was important for institutions to decide on the economic form in addition to selfish motives regarding human nature. The economy's center in a

market society is the exchange, which was small until liberalism increased after the motive of subsistence in classical economics decreased. The communal relationship failed to prevent starvation, and self-interests became prevalent in the market society. This exemplifies the era's employment of economics as a basic rule of society, which Polanyi perceived as problematic. Therefore, Polanyi's thought can lead to a method of creating an institution that controls humanity's self-interests.

Polanyi posits that man is essentially selfish and formally rational at all times. As archaic societies lacked the idea of profit, he did not employ Smith's supposition about primitive man's natural features, or specifically, the alleged predilection for a gainful occupation in addition to rejecting the propensity of exchange. Polanyi considered Smith's idea as correct regarding the relationship between society and the economy. However, he did not accept the propensity of exchange as human nature. Polanyi's critique of economics clarified the erroneous assumption that only by considering the motives of subsistence and gain could exchange create a direct economy without social relationships. This corrected the notion of humankind as individuals only concerned with economic self-interests. Conversely, this did not deny the rational man's existence. Additionally, Polanyi posited that "[f]or the higher the wages, the smaller the inducement to exertion on the part of the native, who unlike the white man was not compelled by his cultural standards to make as much money as he possibly could" (Polanyi 1944: 172–173). This assertion contested the validity of the economic man whose wants remained unfulfilled. If life is assuredly maintained, the subsistence motive does not cause wants to expand infinitely. Similarly, humankind's desire for gain is not unlimited. Therefore, he critiqued economics as a prophetic study on a desk with easy presuppositions of an unrealistic humankind. Additionally, he explored the economy for human subsistence, which functions in an economic society with a foundation in the real world.

## 5 Nineteenth-Century Civilization and Self-interest

Although the previous section discussed humankind as Polanyi presupposed, *TGT* aimed to not only discuss the relationships of man and society with the economy but also to clarify the causes of World Wars and the Great Depression since we can observe an ideal relationship between society and the economy. This section focuses on his interpretation of the world at that time, in which economics becomes a belief. In other words, it is a discussion of his critique on the institution, which is naturally the market society formed by the economic liberalism that relies on individual self-interests.<sup>7</sup>

---

<sup>7</sup>Previous studies on Polanyi usually refer to distinguished and audacious ideas, such as the double movement or the fictitious commodity, as *TGT*'s accomplishment. However, Polanyi's work primarily addressed the origins of world war: "Nineteenth-century civilisation has collapsed. This book is concerned with the political and economic origins of this event, as well as with the great transformation which it ushered in" (Polanyi 1944: 3).

Polanyi referred to capitalistic society as the “nineteenth-century civilization.” He notes four institutional systems that compose this civilization: a balance of power system, the international gold standard, a self-regulating market, and the liberal state. These can be further divided into two ways: domestic-international and political-economic issues. Specifically, the balance of power system is an international and political issue, the international gold standard is international-economic, the self-regulating market is domestic-economic, and the liberal state is domestic-political. He understood the world structure from the perspective of these four elements. The self-regulating market that decides the price and quantity of goods was formed by achieving such a peculiar condition:

Nineteenth-century civilization alone was economic in a different and distinctive sense, for it chose to base itself on a motive only rarely acknowledged as valid in the history of human societies, and certainly never before raised to the level of a justification of action and behavior in everyday life, namely, gain. The self-regulating market system was uniquely derived from this principle. The mechanism, which the motive of gain set in motion, was comparable in effectiveness only to the most violent outbursts of religious fervor in history (Polanyi 1944:31).

The balance of power system functioned as an indispensable factor in forming the nineteenth-century civilization. Although its actual meaning only indicates states' action agendas<sup>8</sup> in international relations, it unintentionally maintains states' independence. Superficially, the balance of power system prevented long-term, devastating wars for a century. However, the balance of power system is fragile, in that the “balance-of-power system could not ensure peace once the world economy on which it rested had failed” (Polanyi 1944: 4–5), as this depends on the global economy. Polanyi's perspective notes international peace based on this economic system. To some extent, the nineteenth-century civilization involved a peaceful situation in which the balance of power regulates some nations.

The self-regulating market is the place in which the motive of gain brings an exchange organized by market patterns where the economy automatically and arbitrarily decides both price and quantity. Naturally, the self-regulating market is the most important factor in Polanyi's critique of the market economy. On the one hand, it produced an unprecedented material welfare as well as the three other institutional systems. On the other hand, it can only exist in rare situations, in that “an institution could not exist for any length of time without annihilating the human and natural substance of society; it would have physically destroyed man and transformed his surroundings” (Polanyi 1944: 3). Essentially, despite the view of the self-regulating market as a utopia, society accepts its features as reality. Polanyi focuses on linking production and sales by noting that “all production is for sale on the market and all incomes derive from such sales” (Polanyi 1944: 72) in the self-regulating market. Moreover, the self-regulating market implies that prices formed when products are

---

<sup>8</sup>In this case, the term “states” indicates European countries, excluding Muslim and Asian states. As Polanyi recognizes, the world has limitations and, thus, he focuses on only Europe; the “nineteenth-century civilization” also means “European civilization.”

sold will dominate, and people cannot avoid this. In other words, the market price becomes the sole dominant criterion.

The international gold standard is a feature of the world economy as organized by specific systems, and this maintains the nineteenth-century civilization as a foundation of the global economic corporative system or the market economy. Specifically, a debacle in the international gold standard system triggered the collapse of the nineteenth-century civilization. The collapse was ultimately triggered by a decrease in the supply of monetary credit to various nations; Polanyi, then, supposes that this depression of such national monetary credit invoked upheaval, and, eventually, war.

Polanyi analyzed international finance such as merchant bankers as the Rothschild family ensured national credit to maintain peace in civilization:

*Haute finance*, an institution *sui generis*, peculiar to the last third of the nineteenth and the first third of the twentieth century, functioned as the main link between the political and the economic organizations of the world. It supplied the instruments for an international peace system, which worked with the help of the Powers, but which the Powers themselves could neither have established nor maintained (Polanyi 1944: 10).

Nations acted based on the motive of gain and acquired profit by loans to each nation to cover war costs. Therefore, wars among nations were preferable for profit. Thus, wars must not be so major as to confuse the monetary market, as they can cause damage by default. Internationally larger bankers preferred small, local, and limited wars. Haute finance-created war, then, led to the collapse of the nineteenth-century civilization. In the view of Polanyi, peacekeeping was a by-product of the international political economy.

Alternately, Polanyi evaluates international finance as having supplied a peacekeeping system, as its key players provided both war and peace. Therefore, each nation could economically connect by market transactions in which an international system of clearance functioned using haute finance actions for profits. Haute finance dislikes major wars not from a moral or ethical point of view but from the motive of gain and acts to deter it. Consequently, a global network of international finance avoided a great war, and prepared a cornerstone to maintain the civilization in collaboration with other nations in Europe:

Their business would be impaired if a general war between the Great Powers should interfere with the monetary foundations of the system. By the logic of facts, it fell to them to maintain the requisites of general peace in the midst of the revolutionary transformation to which the peoples of the planet were subject (Polanyi 1944: 11).

Polanyi insisted that the international gold standard sustained a hundred years' peace through haute financiers' behavior based upon their self-interest. Paradoxically, haute finance's preference for wars created world peace in the civilization.

The liberal state affects nations similar to the balance of power system, in that it identifies the nations that obey economic liberalism. Thus, externally, this state indicates that each nation should act with the same behavior to protect economic liberalism. An example of the liberal state involves nations' free trade. However, these states did not initially insist on free trade, such as in the case of cotton despite the

demand for freedom in its production. Polanyi recognizes that liberalism invariably restricts others, as it does not mean real or perfect freedom. Further, the liberal states stress a prosperous national economy, with interest in domestic issues. Polanyi observed national as well as individual self-interests.

Liberal states connected by the international gold standard system depended on the self-interests of haute finance movements in nineteenth-century civilization. Further, liberal states bring prosperity in the self-regulating market unless political tension expands. Polanyi's idea of nineteenth-century civilization included an international society composed of such nations in which international finance stops peace-sustaining behavior when nations cannot consort with each other.

Polanyi analyzed self-interests through two paths: human nature organizing each economic activity, and the institution or environment surrounding the economy. Moreover, he discussed the self-interests of both the common people and the nation as one integrated subject. Nineteenth-century civilization again prescribes human economic activity in addition to the direct stipulation of economic motives. The economy is an institutional process. By focusing on the collapse of nineteenth-century civilization, the question exists regarding how humanity can address self-interests so as to sustain peace in Polanyi's view. This question can also produce a movement restricting an excessive increase in market transactions in a society (a counter-movement). Alternatively, an inquiry regarding how to confine nations' selfish actions will also arise to internationally and politically connect each nation for peace instead of the international gold standard. Polanyi preferred social policy in the social realm, while he hoped for a political adjustment by such supranational organizations as the League of Nations. As the existence of individual self-interests was not refuted, he aspired to discover a way to handle economic interests with a premise based in the real world.

## 6 Conclusion

This chapter discussed how Polanyi addressed self-interest and economic motives. He considered nineteenth-century civilization, as is unusually compared to other social institutions humanity has experienced. A society is atypical in which the market exchange, which dominates a large part of the economy, depends on the motive of gain—society is not only dominated by the motive of gain, but human activity is also solely based on this motive. Polanyi posits that human nature did not totally convert from the archaic society into modern times, as humanity can naturally alter to any style. Human nature did not change, in that it can modify its actions in response to its surroundings. However, institutional changed sometimes advances self-interests and weakens social ties.

Polanyi believed that humanity has various motives that cause economic activity, and exchanges can naturally be formed from other motives than gain. Additionally, the non-exchange economy exists, which depends on the motive of gain. He rejected the simple supposition created by economic armchair researchers and attempted



to embody substantivism, which perceives human economic activity as an entire truth. He perceived humans as a social existence in economics without denying their self-interests.

Moreover, this chapter indicated self-interest on two levels by discussing nineteenth-century civilization as a starting point of Polanyi's work. In the market economy, in which the economy is embedded into society, international finance's self-interest acts as a peace-maker, in that the preference for small wars prevents substantial catastrophes. Thus, Polanyi's vision leads us to question what those in institutions can create, and not only with a focus on the economy.

Polanyi inquired into the harmony of capitalism with democracy in his late life. He recognized democracy's incompatibility with capitalism based on individual self-interests, in that an ideal policy involves regulating humankind's self-interests on some level without denying them to agree with democracy. He evaluated socialism as nothing more than basic support of a movement that impinges on overheated economic liberalism; he died before the Soviet Union's collapse. However, his socialism aimed not to mainly solve inequality, but to recover economic balance, and the restriction of self-interest does not necessarily conflict with capitalism. One extension of Polanyi's view positively illustrates a world in which non-exchange transactions extends further than at present where society is embedded in the economy by restricting self-interests. Participation in such a market, thus, brings prosperity.

## References

- Baum, Gregory. 1996. *Karl Polanyi on ethics and economics*. Montréal: McGill-Queen's University Press.
- Beckert, Jens. 2009. The great transformation of embeddedness: Karl Polanyi and the new economic sociology. In *Market and society*, ed. Chris Hann and Keith Hart, 38–55. Cambridge: Cambridge University Press.
- Brie, Mark. 2017. *Karl Polanyi in dialogue. A socialist thinker for our times*. Montréal: Black Rose Books.
- Dale, Gareth. 2010. *Karl Polanyi: The limits of the market*. Cambridge: Polity Press.
- Pearson, Harry W. 1977. Editor's introduction. In *The livelihood of man*, ed. Karl, Polanyi, xxv–xxxvi. New York: Academic Press.
- Polanyi, Karl. 1944 [2001]. *The great transformation*. Boston: Beacon Press.
- Polanyi, Karl. 1954. Hamlet. *The Yale Review* 43 (3): 336–350.
- Polanyi, Karl. 1957a [1968]. Aristotle discovers the economy. In *Trade and market in the early empires: Economies in history and theory*, ed. Polanyi, Karl et al., 64–94. Chicago: Gate Way.
- Polanyi, Karl. 1957b [1968]. The economy as instituted process. In *Trade and market in the early empires: Economies in history and theory*, ed. Polanyi, Karl et al., 243–270. Chicago: Gate Way.
- Polanyi, Karl. 1977. *The livelihood of man*. New York: Academic Press.
- Polanyi-Levitt, Kari. 2018. Freedom of action and freedom of thought. In *Karl Polanyi's Vision of a Socialist Transformation*, eds. Michael, Brie and Claus, Thomasberger 18–50. Montréal: Black Rose Books.
- Rogan, Tim. 2017. *The moral economists*. Princeton: Princeton University Press.
- Rotstein, Abraham. 1990. The reality of society: Karl Polanyi's philosophical perspective. In *The life and work of Karl Polanyi*, ed. Kari Polanyi-Levitt, 98–110. Montréal: Black Rose Books.
- Stanfield, James. 1986. *The economic thought of Karl Polanyi*. London: Macmillan Press.

Tandy, David W., and Walter C. Neale. 1994. Karl Polanyi's distinctive approach to social analysis and the case of ancient Greece: Ideas, criticisms, consequences. In *From Political Economy to Anthropology*, ed. Duncan, Colin M. and David W. Tandy, 9–33. Montréal: Black Rose Books.

# Contributions of Science of Free Will to Neuroeconomics and Quantum Decision Theory



Taiki Takahashi

**Abstract** Although thinkers in political science and philosophy such as Thomas Hobbes, John Stuart Mill and Robert Nozick argued that an understanding of “free will” is crucial in considering problems related to democracy and other political institutions, recent advances in the scientific exploration of free will (often associated with rationality and consciousness) in neurobiology, neuroeconomics, and quantum decision theory have largely been ignored in political science. This is also true in recently-developed bio-psychologically-inspired approaches in political science (e.g., political psychology based on evolutionary psychology). In this conceptual review article, we first introduce current findings on the science of free will and then discuss potential implications of these findings for future research in political science. The importance of investigations into future-oriented judgment and decision making (e.g., intertemporal choice) is emphasized.

**Keywords** Free will · Newcomb’s paradox · Neuroeconomics · Neurolaw · Quantum decision theory

## 1 Introduction

Since Thomas Hobbes published his “Leviathan” (Hobbes 1651) in which he denied the existence of free will in usual senses (i.e., free will incompatible with deterministic laws of natural sciences), there have been much controversy concerning the existence of free will in humans in political philosophy. Hobbes was a “*compatibilist*” (i.e., he believed that a certain type of free will which is compatible with deterministic and causal natural laws alone is permitted to exist). If he is correct, it follows that the success of democratic institutions requires the diminishing of the degree of individuals’ “free” choices (the term “free” here is almost equivalent to “self-interested”, but neither equivalent to “non-deterministic” nor “non-causal”).

---

T. Takahashi (✉)

Department of Behavioral Science, Center for Experimental Research in Social Sciences, Research and Education Center for Brain Science, Hokkaido University, Hokkaido, Japan  
e-mail: [ttakahashi@lynx.let.hokudai.ac.jp](mailto:ttakahashi@lynx.let.hokudai.ac.jp)

Natural sciences are generally in harmony with the Hobbesian denial of free will, because concepts of natural laws usually assume physical causality (often associated with determinism) in almost all phenomena in the universe. Recent developments in natural sciences (especially in neurobiology, molecular biology and evolutionary biology) also seem to eliminate the possibilities of the existence of free will in human decisions (Crick 1995; Wegner 2002). Furthermore, in social sciences, the rapidly evolving discipline of “neuroeconomics” which tries to unify economic and neurobiological sciences (Glimcher 2010; Takahashi and Egashira 2013) is also likely to decrease the room for free will. This tendency has started from the beginning of modern political science (and other social sciences such as economics and sociology), since utilitarianism and naturalism are unlikely to be consistent with the concept of (non-deterministic) free will. For instance, Jeremy Bentham (a well-known utilitarian, see Mill and Bentham 1987) thought of the theoretical concept of “human right” as false and/or unintelligible (Russell 1967), for the value of each individual’s right is incompatible with the promotion of social welfare in the society as a whole. Some philosophers such as Immanuel Kant (1998/1785) and Henri Bergson (2013) tried to rescue the concept and ethical value of free will by standing at the philosophical and ethical viewpoints. However, these philosophers’ challenges are not greatly successful in a scientific sense (Russell 1967). Because advances in natural sciences critically influenced the directions of development in social sciences (e.g., Hobbes was seriously influenced by scientific discoveries and mathematical formulations made and developed by Galilei and Descartes in continental Europe), it can reasonably be speculated that modern findings from scientific investigations into free will (conducted in neuroscience, cognitive and behavioral science, decision science, and neuroeconomics) will also affect future developments in political science, although empirical findings and theoretical models in decision sciences (including behavioral economics), so far, have largely been ignored in political science (Wilson 2011; McDermott 2001; McDermott 2004).

This paper consists in the following manner. In Sect. 2, we will introduce the experimental findings on future-oriented decision making in the Newcomb’s paradox (“litmus paper” for the belief in free will, Gardner 2006) and intertemporal choice. Next, in Sect. 3, we will address how these human decisions can be modeled by utilizing theoretical frameworks of neuroeconomics and quantum decision theory. In Sect. 4, several neurobiological studies on free will are introduced. Finally, we will discuss how these theoretical and empirical findings may contribute to future advances in political sciences.

## 2 Future-Oriented Decisions

One of the prominent features of decisions emerged from free will is future-orientation. Free will seems not to be constrained/determined by past causal events. Future orientation is also a characteristics of rational decision making (Hastie and Dawes 2010). Compatibilists such as Thomas Hobbes agree to the latter statement

but disagree to the former. Recent studies in cognitive and behavioral science in addition to neuroeconomics examined future-oriented decisions in an experimental manner.

## 2.1 *Newcomb's Problem*

The libertarian political philosopher Robert Nozick (1969) was the first person to discuss the Newcomb's problem in the context of decision theory (though the thought experiment was first contrived by physicist William Newcomb at Lawrence Livermore Laboratory). Nozick is known for his challenge to Rawlsian principle (1971) of social justice (Nozick 1974). Notice that Nozick's opposition to Rawls' theory of justice is based on anti-utilitarianism.

Let us now see what the Newcomb's problem is. The following version of the Newcomb's problem is adopted (with slight modifications) from mathematical psychologists Shafir and Tversky's experimental work (2004):

[A program developed recently at MIT was applied to analyze the pattern of your previous preferences. Based on that analysis, the program predicted your preference in this problem.

Consider the two boxes. Box A contains 20 points for sure. Box B may or may not contain 250 points. Your options are to:

- (1) Choose both boxes (and collect the points that are in both).
- (2) Choose Box B only (and collect only the points that are in Box B).

If the program predicted, based on observation of your previous preferences, that you will take both boxes, then it left Box B empty. On the other hand, if it predicted that you will take only Box B, then it put 250 points in that box. (So far, the program has been remarkably successful; 92 percent of the participants who chose only Box B found 250 points in it, as opposed to 17 percent of those who chose both boxes.)

To insure that the program does not alter its guess after you have indicated your preference, please indicate to the experimenter in charge whether you prefer both boxes or Box B only. After you indicate your preference, press any key to discover the allocation of points.]

Nozick requested, in his dissertation, us to think about our choice for a few hours or days. A sophisticated behavioral experiment was conducted by Amos Tversky's group to examine which choice is more commonly observed. Shafir and Tversky (2004) reported that most (about 65 percent) people selected option 2 (taking only Box B). However, this choice (one box) by majority is not rational; i.e., it violates Savage's sure-thing principle (Savage 1954): "if we prefer  $x$  to  $y$  given any possible state of the world, then we should prefer  $x$  to  $y$  even when the exact state of the world is not known. (Shafir and Tversky 2004)". Now again consider the Newcomb's problem in terms of Savage's sure-thing principle. There are only two possible states for Box B:

State 1: Box B contains 250 points,

or

State 2: Box B is empty.

In State 1, if you take both boxes, you will win 250 (from Box B) plus 20 (from Box A) points (= 270 points as a total), while if you take Box B only, you will win 250 points. Therefore, you should choose both boxes option (i.e., option 1) for State 1. In State 2, if you take both boxes, you will win 0 (from Box B) plus 20 (from Box A) points (= 20 points as a total), while if you take Box B only, you will win nothing (because Box B is empty in this case). Therefore, you should again choose both boxes option (i.e., option 1). If you follow Savage's sure-thing principle, then you should select option 1, even when you do not know whether Box B is empty or not.

Shafir and Tversky (2004) argues that "there seems to be no defensible rationale for taking just one box." Isaac Asimov, a well-known popular science writer and biochemist, "plumped for the two-box choice as an assertion of his free will"; while Nozick was originally a lukewarm two boxer (Holt 2002). It can thus be said that most people (who are one boxers) make decisions as if they do not have free will and rationality. This choice by the majority is not future-oriented, since the choice of taking just one box is constrained by the prediction of MIT's program made *before* the subjects' decision making (notice that point allocation was determined before the subject's choice), just as addicts are "getting hooked" with their past consumption of harmful drugs. A psychological study by Stanovich and West (1999) on individual differences between one boxers and two boxers revealed that taking both boxes was associated with higher academic performance (better SAT scores) and stronger intellectual curiosity. In order to model majority's "irrational" choice in the Newcomb's problem, a recently-developed framework of quantum decision theory could be adopted. We will introduce this approach in Sect. 3.

## 2.2 *Intertemporal Decision-Making*

Intertemporal choice in decision over time is a type of future-oriented decision making which has extensively been investigated in the field of neuroeconomics (Takahashi 2009). In choice between smaller but sooner reward and larger but later reward, future-oriented subjects prefer larger later reward to smaller sooner one. In contrast, less future-oriented (i.e., impulsive/impatient) subjects prefer sooner but smaller reward. This impulsivity in intertemporal choice has been accounted for by the behavioral/psychological tendency of temporal discounting—devaluation of delayed reward according to an increase in delay until its receipt (Frederick et al. 2002). Intertemporal choice is important also in political decisions, since most policy making is a type of decision over several years.

In behavioral and neuro- economics, various factors have been associated with temporal discounting (see Reynolds 2006, for a review). For instance, heavy smokers

(Bickel et al. 1999; Ohmura et al. 2005) discount delayed reward more rapidly than light and non-smokers, consistent with Becker and Murphy's economic theory of addiction (Becker and Murphy 1988). Less educated and younger subjects also discount delayed reward rapidly (Green et al. 1999, but see Yamane et al. 2013, for a conflicting finding of the effect of age on temporal discounting in Japanese population). Hobbes claimed that deliberation about far future could be associated with anxiety, consistent with our behavioral study of precautionous risk aversion in intertemporal choice (Takahashi et al. 2007). Interestingly, chronic high levels of cortisol (a stress hormone) is related to future-oriented decision in intertemporal choice in men, but not in women (Takahashi 2004; Takahashi et al. 2010).

In neoclassical economics, people's tendency of temporal discounting has been assumed to depend only on the time-interval between delays until the receipt of smaller sooner reward and larger later one. From this assumption, it can be deduced that people's preference between the smaller sooner and the larger later one does not reverse according to mere passage of time (Samuelson 1937; Strotz 1955). However, studies in behavioral economics discovered that people's time preference reverses over time, which is referred to as time-inconsistency in intertemporal choice (Thaler 1981; see Frederick et al. 2002 for a historical development in behavioral economic studies of intertemporal choice). For instance, a subject who prefers [two apples in a month] to [an apple in a week] (i.e., time-interval is three weeks) often also prefers [an apple today] to [two apples in three weeks] (again, the time-interval is three weeks). This intertemporal decision-making pattern is time-inconsistent, in that the time-intervals between the two options are the same (i.e., three weeks) in the two examples. This preference reversal over time is problematic in terms of both theoretically and practically. Theoretically, the time-inconsistency violates the assumption of conventional neoclassical economics and game theory (i.e., time-consistency). Practically, preference reversal may induce various problematic behaviors such as relapse in abstinent addicts and deadline rush (Ainslie 2001). Much work has been done for modeling and elucidating psychological and neural foundation of time-inconsistency, in addition to impatience, in intertemporal choice (see Takahashi 2009, for a review).

"Tempospect theory" generalized the idea proposed by Takahashi (2005) that psychophysics of time perception can account for various "anomalies" in intertemporal choice (Takahashi and Han 2012). We will briefly introduce the roles of psychophysical characteristics of subjective time in accounting for anomalies in intertemporal choice in Sect. 3.2.

As illustrated in the Newcomb's problem and intertemporal choice example, majority of people are not as future-oriented as previously thought. This has several implications for political science. First, the severity of misery in the Hobbesian *state of nature* could be mitigated by the absence of enough future-orientation in human decision making. In the Hobbesian world, future-orientation is tightly connected with maximization of individual's self-interest, resulting in severe conflicts between individuals. Actually, Shafir and Tversky (2004)'s experimental study suggests that non-normative (the term "non-normative" is almost equivalent to "irrational", in this context) choice in the Newcomb's problem (i.e., one box choice) was related to

social cooperation in the Prisoner's Dilemma Game (regarded as "irrational" decision by game theorists, Binmore 2011, but see Goldberg et al. 2005 for further experimental analysis of the relationship between cooperation in prisoner's dilemma and one box choice in the Newcomb's problem), which could solve social dilemma. Next, accumulation of empirical findings may promote people's mechanistic view on human decision and behavior, which may eventually reduce the respect for liberty and freedom of individuals in political leader's policy making. This is not problematic for political philosophers such as Hobbes and Bentham, but problematic for classical liberals such as Mill (2008) and Becker (2007) and more confident libertarians such as Nozick. Proposing the answer to this reservation is beyond the scope of the present article but it is to be noted that this tendency is an example showing that advances in scientific knowledge often discount ethical and social value of individual's free choice.

### 3 Models in Neuroeconomics and Quantum Decision Theory

In order to model and account for anomalies in human decision making, recent studies try to unify decision theory with neurobiology ("neuroeconomics") or quantum probability theory ("quantum decision theory"). Neuroeconomics is a reductionistic approach, while quantum decision theory is a non-reductionistic approach. It should probably be now explained why quantum decision theory is a non-reductionistic approach, for quantum physics describes microscopic, subatomic physical phenomena. A brief answer to this question is that quantum decision theory does not try to model human decision making at the subatomic level, but utilize mathematical characteristics of quantum probability for modelling non-normative human judgment and decision making (Busemeyer and Bruza 2012; Khrennikov 2010; Yukalov and Sornette 2011; Takahashi 2014). In contrast, neuroeconomic theory has started to analyze anomalies in economic decision making at the molecular and cellular levels (Takahashi 2009, 2010a, b).

#### 3.1 *Quantum Decision Theory*

In usual probability theory (i.e., Kolmogorovian probability theory), the following formula (a law of total probability, abbreviated as LTP hereafter) on the relation between unconditional and conditional probabilities should be satisfied:

$$P(A) = P(B)P(A|B) + P(\text{not}B)P(A|\text{not}B) \quad (1)$$



where  $P(A)$  denotes the probability that event  $A$  occurs,  $P(B)$ , the probability that event  $B$  occurs, and  $P(\text{not}B)$ , the probability that event  $B$  does not occur.  $P(A|B)$  is the conditional probability that the probability that event  $A$  occurs under the condition that event  $B$  occurred. Similarly,  $P(A|\text{not}B)$  is the conditional probability that the probability that event  $A$  occurs under the condition that event  $B$  did not occur. Let us apply this formula (LTP) to the situation of the Newcomb's problem. Then, event  $A$  corresponds to the selection of both boxes, event  $B$  corresponds to the state that Box  $B$  contains 250 points:

$$P(\text{twoboxchoice}) = P(B = 250)P(\text{twoboxchoice}|B = 250) + P(B = 0)P(\text{twoboxchoice}|B = 0) \quad (2)$$

As we considered above, if Box  $B$  contains 250 points, we surely should choose two boxes (i.e.,  $P(\text{two box choice}|B = 250) = 1$ ) and even if Box  $B$  is empty, we again surely should choose two boxes (i.e.,  $P(\text{two box choice}|B = 0) = 1$ ). Therefore, Eq. 2 is then

$$P(\text{two box choice}) = P(B = 250) + P(B = 0) = 1 \quad (3)$$

because Box  $B$  has only two possible states: either empty or containing 250 points. The readers may be aware that Eq. 3 corresponds to Savage's sure-thing principle. It can hence be said that one-boxers in the Newcomb's problem violates LTP (and sure-thing principle is a simple special case of LTP. In other words, one-boxers in the Newcomb's problem violate this fundamental law of Kolmogorovian probability theory.

In quantum probability theory, LTP could be modified under some assumptions (Franco 2009; Khrennikov 2010) in the following manner:

$$P(A) = P(B)P(A|B) + P(\text{not}B)P(A|\text{not}B) + (\text{quantum interference term}) \quad (4)$$

due to non-classical nature of quantum theory (for more complete quantum formalism, see Cheon and Takahashi 2010 and Cheon and Takahashi 2012). The quantum interference term (the additional last term in Eq. 4) is either negative or positive, relaxing the classical LTP (Eq. 1). We can then proceed to utilize quantum-generalized law of total probability (Eq. 4) to model violation of Savage's sure-thing principle in the Newcomb's problem by the majority of human decision makers. Some philosophers argued that because quantum theory has a nature of indeterminacy, quantum theory can possibly be a scientific foundation of free will in considering the relationship between neurobiology of free will and political power (Searle 2006). But, in the case of Newcomb's problem, our present analysis suggests the opposite: majority's choice (one box), apparently lacking future-orientated free will, is described with quantum probabilistic decision model; while future-oriented normative decision is described with classical (i.e., non-quantum) probability theory. Interestingly, quantum theoretical logic allows the existence of democracy (Segre

2008) without inconsistency shown by Kenneth Arrow's impossibility theorem on democracy. Also, quantum decision theory has started to analyze US political party governance (Khrennikova et al. 2014). These developments in quantum decision theory implies that political science will finally be influenced by findings from human decision making models based on quantum theoretical formalism.

### 3.2 Neuroeconomic Theory of Intertemporal Choice

In order to capture essential features of human decision over time, the following q-exponential temporal discounting model has been utilized in neuroeconomics and econophysics (Cajueiro 2006; Takahashi 2009):

$$V_q(t) = \frac{V_q(0)}{(1 + k_q(1 - q)t)^{\frac{1}{1-q}}} \quad (5)$$

where  $k_q$  and  $q$  are free parameters indicating impulsivity (impatience) at  $t = 0$  ( $t$  is a delay until the receipt of the delayed reward) and deviation from neoclassical rationality in intertemporal choice, respectively. Larger  $k_q$  values represent greater temporal discounting (impulsivity) at  $t = 0$ . If  $q = 1$ , Eq. 5 corresponds to the exponential discount function originally proposed in neoclassical economics (Samuelson 1937):

$$V_1(t) = V_1(0)\exp(-k_1t) \quad (6)$$

If  $q = 0$ , the time discount function is the following (simple) hyperbolic function (Mazur 1987):

$$V_0(t) = \frac{V_0(0)}{1 + k_0t} \quad (7)$$

which has often been utilized in behavioral economics (Ainslie 2001; Frederick et al. 2002). Notice that perfect time consistency exist only in Eq. 6 (exponential discounting, see Takahashi 2009 and Takahashi 2011 for mathematical proofs). The next question may be: what causes human time-inconsistent intertemporal decision-making, instead of time-consistent exponential time discounting? Takahashi (2005) tried to answer this question. The idea is that *exponential discounting with logarithmic psychological time* induces the q-exponential time discount function in physical (objective) time. From this theory, it is predicted that (i) temporal discounting is exponential in psychological time, and (ii) modulation of psychological time via alteration of neural activities mediating psychological time will change temporal discounting in objective time. Prediction (i) was confirmed by behavioral and psychophysical experiments (Zauberman et al. 2009; Han and Takahashi 2012). Prediction (ii) is

consistent with our recent neurogenetic study which demonstrated that the genetic polymorphism (i.e., individual differences in a gene) in dopamine D2 receptor (modulating time perception, Wiener et al. 2011) is associated with temporal discounting (Kawamura et al. 2013). Also, depressed patients (known to have reduced serotonergic activities) were shown to have strong temporal discounting and reduced time-consistency in intertemporal choice (Takahashi et al. 2008). Furthermore, Takahashi (2007) experimentally demonstrated that intertemporal choice for other is more impulsive and time-inconsistent in comparison to intertemporal choice for self, indicating that paternalism may not work well in the problem of intertemporal decision making. This is a prominent finding suggesting that respect for individual's own decision is important for the problem of decision over time.

Recently, it has been speculated that (seemingly) quantum theoretical human decision making may be due to nonlinear psychophysical effect of neural networks on subjectively-perceived quantities (Takahashi and Cheon 2012). This speculation implies that quantum decision theory will finally be unified with the findings in neurobiology, just as economic decision theory has been combined (and is now being unified) with neuroscience in the discipline of neuroeconomics within a recent decade.

## 4 Neurobiology of Free Will

Neurobiological explorations for mechanisms of free will have been conducted relatively independently of neuroeconomics and quantum decision theory. The first serious electrophysiological studies (i.e., neurophysiological studies with electrical measurements of brain waves) of free will was conducted by Benjamin Libet and his colleagues (Libet et al. 1982, 1983). In his experiment, it was shown that the perceived timing of free will-guided decision making (intentional decision) is *later* than the onset of electrical activity (referred to as “readiness potential”) in brain's motor controlling area. The naïve interpretation of the Libet's experiment is that free will-guided decision processes initiates *after* (not before) unconscious body movement preparation occurs. It has therefore been claimed that Libet's experimental result is scientific evidence of the absence of non-deterministic and non-causal (i.e., independent from past events) free will. Modern non-invasive neuroimaging techniques (e.g., functional magnetic resonance imaging, Ogawa 1991) have been combined with Libet-type behavioral paradigms to ask the question: what parts of the brain are responsible for free will (i.e., voluntary and intentional decision making)? According to many neuroimaging and brain lesion studies, the anterior cingulate cortex and the supplementary and pre motor areas have been associated with voluntary decision making of movement (see Kriehoff et al. 2011, for a review of current findings and remaining controversies). Interestingly, it was reported that free will-guided (intentional) and externally guided (unintentional) actions differ in the *perception* of time interval between the timing of initiating actions and the timing of the occurrence of the consequence of the actions (Haggard et al. 2002). This study suggests a possible

(but largely understudied) connection between free will and time-based decision such as intertemporal choice which utilizes psychological time (Lucci 2013). This potential link is interesting, for impulsivity in intertemporal choice is reportedly related to a decrease in perceived continuity between future and present selves (Frederick 2003; Hershfield 2011).

## 5 Potential Contributions of Science of Free Will to Political Science

Historically, the influences of the developments in scientific knowledge on social sciences, both in principle and in practice, promoted utilitarian views and reduced beliefs in free will, at the cost of respect for each individual's freedom and liberty. Utilitarian views successfully mitigated the severity of punishment inflicted upon criminals via enhancement of mechanistic and deterministic view on criminals' misconduct, in comparison to criminal justice systems based on religious ethics. Actually, a recent experimental study on punishing behavior in neurolaw (a discipline trying to unify legal studies and neuroscience, Jones et al. 2013; Takahashi 2012) demonstrated that mechanistic view on human behavior reduces punishing behavior (Shariff et al. 2014), consistent with the supposed relationship between scientific determinism and reduced punishment. Utilitarians put heavy weight on the prevention of criminal behavior, rather than imposition of cruel punishment on sinners (remember that Hobbes argued that political power is legitimate due *solely* to its ability to protect individuals' peaceful lives by preventing conflicts between them, not to its sacredness). In this regard, enforcement of utilitarianism had, as a whole, better influences on human welfare. But, is there any threshold for utilitarianism to do more harm than good? Robert Nozick was strongly against utilitarian ethics in defense of individual's value. As introduced earlier, his speculation on free will with the adoption of thought experiment on the Newcomb's problem, has been started to be examined in an empirical manner. It could be proven that non-utilitarian political philosophy is not incompatible with scientific knowledge.

A neuroeconomic study (Aoki et al. 2014) reported that valuation system in the brain (the ventromedial prefrontal cortex) encode the value of equality in choice opportunities (rather than choice outcomes) between individuals. Because utilitarianism is a type of consequentialism, it is not completely unlikely that neuro-behavioral evidence for the existence of human's non-consequential judgment and decision making imposes a maximal limit on the degree of consequential utilitarianism even in naturalistic ethics. For instance, our study showed that most non-criminal citizens' punishing behavior is socially-inefficient in terms of consequential utilitarianism (Yokoyama and Takahashi 2013). Whether this inefficient punishing behavior should be "de-biased" by paternalistic governmental interventions or not ought to be analyzed with the knowledge of more detailed neurobiological understanding of human punishment behavior. Collectively, future studies in political science (and

political philosophy) may seriously be affected by deeper understanding of neurobiological and biophysical foundations of human judgment and decision making, in addition to evolutionary foundations of political psychology (McDermott et al. 2008).

## References

- Ainslie, G. 2001. *Breakdown of will*. Cambridge: Cambridge University Press.
- Aoki, R., M., Matsumoto, Y., Yomogida, K., Izuma, K., Murayama, A., Sugiura, C.F., Camerer, R. Adolphs and K. Matsumoto. 2019. Social equality in the number of choice options is represented in the ventromedial prefrontal cortex. *Journal of Neurosciences* 34 (18): 6413–6421.
- Becker G.S. 2007. Libertarian paternalism: A critique. <http://www.becker-posner-blog.com/2007/01/libertarian-paternalism-a-critique-becker.html>.
- Becker, G.S., and K.M. Murphy. 1998. A theory of rational addiction. *Journal of Political Economy* 96(4): 675–700.
- Bergson, H. 2013. *Time and free will, an essay on the immediate data of consciousness*. Hard Press Publishing.
- Bickel, W.K., A.L. Odum, and G.J. Madden. 1999. Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology (Berlin)*. 146 (4): 447–454.
- Binmore, K. 2011. *Rational decisions*. Princeton University Press.
- Busemeyer, J.R., and P. Bruza. 2012. *Quantum models of cognition and decision*. Cambridge University Press.
- Cajueiro, D.O. 2006. A note on the relevance of the q-exponential function in the context of intertemporal choices. *Physica A: Statistical Mechanics and Its Applications* 364: 385–388.
- Cheon, T., and T. Takahashi. 2010. Interference and inequality in quantum decision theory. *Physics Letters A* 375 (2): 100–104.
- Cheon, T., and T. Takahashi. 2012. Quantum phenomenology of conjunction fallacy. *Journal of the Physical Society of Japan* 81 (10): 104801.
- Crick, F. 1995. *The astonishing hypothesis: The scientific search for the soul*. Scribner reprint edition. 1995.
- Franco, R. 2009. The conjunctive fallacy and interference effects. *Journal of Mathematical Psychology* 53 (5): 415–422.
- Frederick, S. 2003. *Personal Identity*, 89. Time and decision: Economic and psychological perspectives of intertemporal choice.
- Frederick, S., G. Loewenstein, and T. O'Donoghue. 2002. Time discounting and time preference: A critical review. *Journal of Economic Literature* 40: 351–401.
- Gardner, M. (2006) aha! A two volume collection: aha! Gotcha aha! Insight (The Mathematical Association of America).
- Green, L., J. Myerson, and P. Ostraszewski. 1999. Discounting of delayed rewards across the life span: Age differences in individual discounting functions. *Behavioural Processes* 46: 89–96.
- Glimcher, P.W. 2010. *Foundations of Neuroeconomic Analysis*. USA: Oxford University Press.
- Goldberg, J., L. Markoczy, and G.L. Zahn. 2005. Symmetry and the illusion of control as bases for cooperative behavior. *Rationality and Society* 17 (2): 243–270.
- Haggard, P., S. Clark, and J. Kalogeras. 2002. Voluntary action and conscious awareness. *Nature Neuroscience* 5: 382–385.
- Han, R., and T. Takahashi. 2012. Psychophysics of time perception and valuation in temporal discounting of gain and loss. *Physica A: Statistical Mechanics and Its Applications* 391 (24): 6568–6576.
- Hastie, R., and R.M. Dawes. 2010. *Rational choice in an uncertain world*, 2nd ed. Thousand Oaks, CA: Sage.

- Hershfield, H.E. 2011. Future self-continuity: How conceptions of the future self transform intertemporal choice. *Annals of the New York Academy of Sciences* 1235 (1): 30–43.
- Hobbes, T. 1651. Leviathan or the matter, forme, & power of a common wealth ecclesiastical and civil. <http://www.gutenberg.org/files/3207/3207-h/3207-h.htm>.
- Holt, J. 2002. Thinking inside the boxes. *Slate*. [http://www.slate.com/articles/arts/egghead/2002/02/thinkinginside\\_the\\_boxes.html](http://www.slate.com/articles/arts/egghead/2002/02/thinkinginside_the_boxes.html).
- Jones, O.D., A.D. Wagner, D.L. Faigman, and M.E. Raichle. 2013. Neuroscientists in court. *Nature Reviews Neuroscience* 14 (10): 730–736.
- Kant, I. 1998. *Grundlegung zur Metaphysik der Sitten [The Ground work of the Metaphysics of Morals]*, trans. and ed. M. Gregor. New York, NY: Cambridge University Press. (Original work published 1785).
- Kawamura, Y., T. Takahashi, X. Liu, N. Nishida, Y. Noda, A. Yoshikawa, T. Umegaki, and T. Sasaki. 2013. Variation in the DRD2 gene affects impulsivity in intertemporal choice. *Open Journal of Psychiatry* 3: 26–31.
- Khrennikov, A.Y. 2010. Ubiquitous quantum structure: From psychology to finance.
- Khrennikova, P., E. Haven, and A. Khrennikov. 2014. An application of the theory of open quantum systems to model the dynamics of party governance in the US political system. *International Journal of Theoretical Physics* 53 (4): 1346–1360.
- Kriehoff, V., F. Waszak, W. Prinz, and M. Brass. 2011. Neural and behavioral correlates of intentional actions. *Neuropsychologia* 49: 767–776.
- Libet, B., E.W. Wright Jr., and C.A. Gleason. 1982. Readiness-potentials preceding unrestricted ‘spontaneous’ vs. pre-planned voluntary acts. *Electroencephalography Clinical Neurophysiology* 54(3): 322–335.
- Libet, B., C.A., Gleason, E.W. Wright, and D.K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 106 (Pt 3): 623–642.
- Lucci, C.R. 2013. Time, self, and intertemporal choice. *Frontiers Neurosciences* 27 (7): 40.
- Mazur, J.E. 1987. An adjusting procedure for studying delayed reinforcement. In *Quantitative analysis of behavior: The effects of delay and of intervening events on reinforcement value*, vol. 5. Ed. M.L. Commons, J.E. Mazur, J.A. Nevin, and H. Rachlin, 55–73. NJ: Hillsdale.
- McDermott, R. 2001. The psychological ideas of Amos Tversky and their relevance for political science. *Journal of Theoretical Politics* 13: 5–33.
- McDermott, R. 2004. Prospect theory in political science: Gains and losses from the first decade. *Political Psychology* 25: 289–312.
- McDermott, R., J.H. Fowler, and O. Smirnov. 2008. On the evolutionary origin of prospect theory preferences. *The Journal of Politics* 70 (02): 335–350.
- Mill, J.S. 2008. *On liberty and other essays (Oxford World’s Classics)*. Oxford: Oxford University Press.
- Mill, J.S., and J. Bentham. 1987. *Utilitarianism and other essays*, ed. A. Ryan. New York, NY: Penguin Books.
- Nozick, R. 1969. Newcomb’s problem and two principles of choice. In *Essays in honour of Carl G. Hempel*, ed. N. Rescher, 114–146. Dordrecht, the Netherlands: Reidel.
- Nozick, R. 1974. *Anarchy, state, and utopia*. New York: Basic Books.
- Ogawa, S., T.M. Lee, A.R. Kay, and D.W. Tank. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America* 87 (24): 9868–9872.
- Ohmura, Y., T. Takahashi, and N. Kitamura. 2005. Discounting delayed and probabilistic monetary gains and losses by smokers of cigarettes. *Psychopharmacology (Berlin)* 182: 508–515.
- Rawls, J. 1971. *A theory of justice*. Cambridge, MA: Harvard University Press.
- Reynolds, B. 2006. A review of delay-discounting research with humans: relations to drug use and gambling. *Behavioural Pharmacology* 17: 651–667.
- Russell, B. 1967. *A history of western philosophy*. Simon & Schuster/Touchstone (October 30, 1967).

- Samuelson, P.A. 1937. A note on measurement of utility. *Review of Economic Studies* 4: 155–161.
- Savage, L.J. 1954. *The foundations of statistics*. New York: Wiley & Sons.
- Searle J.R. 2006. *Freedom and neurobiology: Reflections on free will, language, and political power*. Columbia University Press.
- Segre, G. 2008. Quantum democracy is possible. arXiv preprint [arXiv:0806.3667](https://arxiv.org/abs/0806.3667).
- Shafir, E., and A. Tversky. 2004. Thinking through uncertainty: Non-consequential reasoning and choice. In *Preferences, belief, and similarity*, ed. E. Shafir, 703–727. Cambridge, MA: MIT Press.
- Shariff, A.F., J.D. Greene, J.C. Karremans, J.B. Luguri, C.J. Clark, J.W. Schooler, R.F. Baumeister, and K.D. Vohs. 2014. Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science* 25 (8): 1563–1570.
- Stanovich, K.E., and R.F. West. 1999. Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology* 38: 349–385.
- Strotz, R.H. 1955–56. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies* 23(3): 165–180.
- Takahashi, T. 2004. Cortisol levels and time-discounting of monetary gain in humans. *NeuroReport* 15: 2145–2147.
- Takahashi, T. 2005. Loss of self-control in intertemporal choice may be attributable to logarithmic time-perception. *Medical Hypotheses* 65 (4): 691–693.
- Takahashi, T. 2007. A comparison of intertemporal choices for oneself versus someone else based on Tsallis' statistics. *Physica A: Statistical Mechanics and Its Applications* 385 (2): 637–644.
- Takahashi, T. 2009. Theoretical frameworks for neuroeconomics of intertemporal choice. *Journal of Neuroscience, Psychology, and Economics*. 2: 75–90.
- Takahashi, T. 2010a. Toward molecular neuroeconomics of obesity. *Medical Hypotheses* 75 (4): 393–396.
- Takahashi, T. 2010b. A neuroeconomic theory of bidirectional synaptic plasticity and addiction. *Medical Hypotheses* 75 (4): 356–358.
- Takahashi, 2011. A neuroeconomic theory of rational addiction and nonlinear time-perception. *Neuroendocrinology Letters* 32: 221–225.
- Takahashi, T. 2012. Molecular neuroeconomics of crime and punishment: Implications for neurolaw. *Neuroendocrinology Letters* 33 (7): 667–673.
- Takahashi, T. 2014. Toward a physical theory of quantum cognition. *Topics in cognitive science* 6 (1): 104–107.
- Takahashi, T., K. Ikeda, and T. Hasegawa. 2007. A hyperbolic decay of subjective probability of obtaining delayed rewards. *Behavioral and Brain Functions* 3: 52.
- Takahashi, T., H. Oono, T. Inoue, S. Boku, Y. Kako, Y. Kitaichi, I. Kusumi, T. Masui, S. Nakagawa, K. Suzuki, T. Tanaka, T. Koyama, and M.H. Radford. 2008. Depressive patients are more impulsive and inconsistent in intertemporal choice behavior for monetary gain and loss than healthy subjects—an analysis based on Tsallis' statistics. *Neuroendocrinology Letters* 29 (3): 351–358.
- Takahashi, T., M. Shinada, K. Inukai, S. Tanida, C. Takahashi, N. Mifune, H. Takagishi, Y. Horita, H. Hashimoto, K. Yokota, T. Kameda, and T. Yamagishi. 2010. Stress hormones predict hyperbolic time-discount rates six months later in adults. *Neuroendocrinology Letters* 31: 616–621.
- Takahashi, T., and T. Cheon. 2012. A nonlinear neural population coding theory of quantum cognition and decision making. *World Journal of Neuroscience* 2: 183–186.
- Takahashi, T., and R. Han. 2012. Tempospect theory of intertemporal choice. *Psychology* 3: 555–557.
- Takahashi, T., and S. Egashira. 2013. Hayek's sensory order, Gestalt neuroeconomics, and quantum psychophysics. In *Hayek and behavioral economics (Archival Insights Into the Evolution of Economics)*, ed. Smith, Vernon L. (Prefis), Frantz, Roger, and Robert Leeson. Palgrave Macmillan.
- Thaler R.H. 1981. Some empirical evidence on dynamic inconsistency. *Economic Letters*: 201–207.
- Wegner, D.M. 2002. *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wiener, M., F.W. Lohoff, and H.B. Coslett. 2011. Double dissociation of dopamine genes and timing in humans. *Journal of Cognitive Neuroscience* 23 (10): 2811–2821.

- Wilson, R.K. 2011. The contribution of behavioral economics to political science. *Annual Review of Political Science* 14: 201–223.
- Yamane, S., T. Takahashi, A. Kamesaka, Y. Tsutsui, and F. Ohtake. 2013. Socio-emotional status, education, and time-discounting in japanese non-smoking population: A multi-generational study. *Psychology* 4 (2): 124–132.
- Yokoyama, T., and T. Takahashi. 2013. Mathematical neurolaw of crime and punishment: The q-exponential punishment function. *Applied Mathematics* 4: 1371–1375.
- Yukalov, V.I., and D. Sornette. 2011. Decision theory with prospect interference and entanglement. *Theory and Decision* 70 (3): 283–328.
- Zauberman, Gal, B. Kyu Kim, Selin A. Malkoc, and James R. Bettman. 2009. Discounting time and time discounting: Subjective time perception and intertemporal preferences. *Journal of Marketing Research* 46 (4): 543–556.



# An Institution to Reconcile Self-interests: Hayek on the Notion of the Economic Agent and the Establishment of Appropriate Rules



Masanori Taishido

**Abstract** It is commonly accepted that Hayek's economic philosophy is based on the concept of social prosperity resulting from the selfish behaviour of individuals (a kind of consequentialism). However, he did not simply praise individual egoism or advocate laissez-faire based on it. Although the subjective motivations of individuals can include a variety of things, they do not directly create a society. Instead, society emerges as 'unintended consequences' of the interaction of subjective behaviours ('spontaneous order'). In this sense, Hayek's economic philosophy is unique in that it presupposes a variety of individual motives that are different from simple utility maximisation behaviour, while at the same time criticising reductionist social construction, including utilitarianism. On the other hand, it should be noted that Hayek harshly criticised the phenomenon in which individual optimisation actions that deviate from appropriate rules often become a political collective interest that distort the market framework. In particular, he has always warned that the motivations of individuals to defend their vested rights tend to collude in the name of 'social justice' and transform into 'group selfishness', thus distorting the framework of market and social order. Therefore, this chapter clarifies Hayek's advocacy of his own institutional design as a preventive measure. Hayek's theory of rules is designed to protect the diverse interests and freedoms of individuals from the threat of collective and uniform interests.

**Keywords** Group selfishness · Dispersed knowledge · The rule of law · Empirical natural law · Republicanism

---

M. Taishido (✉)  
Department of Economics, Toyo University, Tokyo, Japan  
e-mail: [taishido@toyo.jp](mailto:taishido@toyo.jp)

## 1 Introduction

This chapter aims to examine the image of an economic agent via Friedrich Hayek's view of self-interest. It is commonly accepted that Hayek's economic thought emphasises the significance of property rights based on the self-interested actions of individuals and a social prosperity concept that assumes that a market economy is an intermediary between individuals and society. For this reason, Hayek praised the contributions of British thinkers, such as Bernard Mandeville and David Hume as precursors of his ideas (Hayek 1949). He basically extolled individualistic subjectivism, but was not a proponent of the *laissez-faire* theory, which only assumes such egoism and individual property rights. As we discuss in Sect. 2, Hayek did not make an issue of whether the motivations for human actions were self-centred. The main point of his argument was that in either case (even with the most noble altruist), individual understanding is only a small part of social understanding and the things that can be accomplished by an individual are relatively limited because knowledge is dispersed. According to him, an individual's subjective motivations are both self-centred and altruistic; however, these motivations do not directly create society. Hayek harshly criticised this type of reductionist attitude, terming it 'constructivism'. He considered such constructivism to be represented in Benthamite utilitarianism and movements to build ideal, altruism-based societies under the banner of 'social justice' (that is, socialism). To Hayek, the appearance of society as a 'spontaneous order' emergent as an 'unintended consequence' of individuals' subjective actions is of utmost importance. Thus, Hayek assumed diverse individual motivations that are different from the simple utility maximisation behaviours of mainstream economics. In fact, Hayek consistently emphasised the dangers of individual-optimising behaviour based on such motivations, suggesting that they morph into political 'group self-interest' and distort the framework of the market (Hayek 1979, Chap. 15).

The second objective of this chapter is to discuss Hayek's advocacy of his own system design as a countermeasure to this distortion. In Sect. 3, we discuss Hayek's warning of the dangers of 'group self-interest'. Hayek developed his idea of 'spontaneous order', suggesting that free actions in a market based on the motivations of individuals generate the greatest societal consequences, even though this is not achieved by simple *laissez-faire*. Rather, Hayek continually sounded the warning that the motivations of individuals trying to defend vested rights may degenerate into 'group self-interest', consequently distorting the market framework by colluding under the guise of 'social justice', thus he harshly criticised Benthamite utilitarianism and its attempt to create an ideal society by aggregating the utility of individuals. Although most human action generally stems from so-called self-interest, Hayek felt that it was impossible to quantify or assess self-interest by any type of measurement. In particular, he suggested that motivations, although able to shape society, are dispersed among individuals and cannot be directly extracted, explicated, and/or used for social policy. In addition, if we suppose that a liberal society is not a survival-of-the-fittest 'war of all against all', then a social system that guides individual action toward a stable societal order is necessary.

In Sect. 4, we discuss that to this end, Hayek emphasised the social importance of adhering to rules, making his own propositions on legislative institutions. By following general rules (nomos) that explicitly forbid particular things, individuals adopt various strategies and are able to exist in harmony with one another. In the event that they do clash, and the situation worsens into a dispute, it is important to have judicial organs that create new rules to mediate those disputes. Establishing a liberal order requires a framework of rules amenable to continually making such revisions. Thus, Hayek developed his own liberalism based on the concept of the ‘rule of law’, although this view is not unique to Hayek. This style of liberal thinking, including libertarianism, is characterised by ‘the priority of the right over the good’ (Sandel 1984). Rules of justice are elevated to the position of superior concepts so that the ‘conceptions of good’—as embodied in the values of individuals and groups—do not conflict with one another and be mediated, should they clash. Therefore, the issue here is the method by which rules are made legitimate or derived, and in that sense, Hayek’s argument occupies a unique space even among many liberal thinkers.

Especially we focus on the relationship between Hayek’s theory of the rule of law and the perspective on natural law and republicanism. The tradition of natural law is one of the starting points of economics and has changed from the time of ancient Greece following the identification of the basis of law and the manner in which it secures the stability of society. Extant research on Hayek’s legal theory shows that it has a hierarchical structure; that is, it is a secular and empirical natural law centred on ‘opinions’ (See Note 12). Needless to say, there is a deep connection between liberalism and democracy. However, we had seen ‘bargaining’ democracies as typical examples of conflict between self-interest and vested rights, majority votes as sometimes inviting despotism, or perhaps socialist or fascist regimes being brought into power in the name of democracy in twentieth century. Likewise, in our current age, we often see the phenomena of liberalism and democracy in conflict, particularly concerning the possibility of democracy spelling doom for liberalism. As we see in Sect. 4.2, some recent research (Pocock 1975; Pettit 1997; Skinner 1998) on republicanism suggests that it cannot be reduced to either liberalism or democracy, but that it exists dually as an undercurrent in both and a ‘mediator’ between them. Hayek tried to reduce the political arena as much as possible and praised the free actions of individuals in the market, but at the same time, his argument is largely related to a kind of republicanism. In Hayek’s perspective, totalitarianism (as manifested in fascism and socialism) is ‘atavism’ toward ancient primitive and communal feelings that seek lapsed ties. Hayek particularly noted many times, as mentioned before, that the optimising behaviour of people rent-seeking and trying to acquire vested rights that are not based on general rules becomes ‘group self-interest’, which further carries the risk of distorting the framework of social order in the name of ‘social justice’. Hayek stated the need for a framework of proper legislation to prevent this and to protect the interests of diverse individuals from the threat of collective or uniform interests.

## 2 Hayek's View of Economic Agents and Economic Order as a Catallaxy

### 2.1 Hayek's View of Economic Agents

It is commonly accepted that Hayek's libertarianism is built on individualism, although he does not deny the existence of altruistic elements in the human spirit, whether innate or acquired. Rather, Hayek consistently maintained that altruism is one of the factors of human nature that can especially be seen in primitive societies. In fact, this argument remains between Hayek's *'The Road of Serfdom'* (1944) and *'Individualism and Economic Order'* (1949) through *'The Constitution of Liberty'* (1960) and *'Law, Legislation and Liberty'* (1973–1979).<sup>1</sup> For example:

For our problem it is of minor importance whether the ends for which any person cares comprehend only his own individual needs, or whether they include the needs of his closer or even those of his more distant fellows—that is, whether he is egoistic or altruistic in the ordinary senses of these words. The point which is so important is the basic fact that it is impossible for any man to survey more than a limited field, to be aware of the urgency of more than a limited number of needs (Hayek 1944, p. 66. In addition, see also Hayek 1949, p. 14; Hayek 1978b, p. 252, 1979, p. 70; etc.).

Hayek continually emphasised the importance of 'dispersed knowledge' in individuals that cannot be uniformly used by a single political agent to design society. Moreover, while noting that the preferences of economic agents, whether self-centred or altruistic, are subjective, he also argued that their societal impact when constructing subjectivity is not excluded; rather, the preferences assume the subjectivity or individualism that Hayek asserted.<sup>2</sup> Further, Hayek noted:

---

<sup>1</sup>It has been noted by multiple theorists that Hayek's thought changed direction several times over the years. For example, Vanberg (1986) noted that Hayek's thought moved from methodological individualism to methodological holism, and that these are incompatible in nature and fragmented. In addition, Fleetwood (1995) asserted that from the perspective of critical realism, Hayek's thought could be divided into three periods: Hayek I, strong empiricism in nature; Hayek II, strong subjectivism in nature; and Hayek III, strong realism in nature. These issues were generalised by Caldwell (1988, 2004) in 'Hayek's Transformation' and are important for researchers—although we do not deal with them further here.

<sup>2</sup>Hayek criticised the image of a human as a 'homo economicus' who is always engaged in strictly rational behaviour, as imagined in neoclassical economics, and alluded to the arguments of Smith in the following:

It would be nearer the truth to say that in their (Adam Smith and his group: added by the quoter) view, man was by nature lazy and indolent, improvident and wasteful, and that it was only by the force of circumstances that he could be made to behave economically or carefully to adjust his means to his ends (Hayek 1949, p. 11)

Further, as can be seen in the criticism of Galbraith's concept of 'dependence effect' (Hayek 1961), this does not remove the possibility of others' influence in subjective formation, and instead, it is predicated on this sort of sociality of individuals.

It is, however, a misunderstanding to represent this as an effort to make 'economic ends prevail over others. There are, in the last resort, no economic ends. The economic efforts of the individuals as well as the services, which the market order renders to them, consist in an allocation of means for the competing ultimate purposes, which are always non-economic. The task of all economic activity is to reconcile the competing ends by deciding for which of them the limited means are to be used. The market order reconciles the claim of the different non-economic ends by the only known process that benefits all (emphasis mine: Hayek 1976, p. 113)

Thus, real economic agents differ from the so-called homo economicus and do not pursue purely economic objectives. The original 'objectives' of individuals achieved by economic action might involve both 'non-economic' and 'altruistic' goals. In the first place, society has no objectives that are purely 'economic'; these are nothing more than the means for 'non-economic' values and objectives. Moreover, objectives achieved by economic action are not at all problematic even when they are subjectively 'altruistic', as long as one voluntarily bears the costs. Simultaneously, these objectives lose their validity when sacrificing others and imposing themselves in the name of 'social justice' as discussed in the next section.

Only where we ourselves are responsible for our own interests and are free to sacrifice them has our decision moral value. We are neither entitled to be unselfish at someone else's expense nor is there any merit in being unselfish if we have no choice. The members of a society who in all respects are made to do the good thing have no title to praise (Hayek 1944, p. 231).

In the end, the element comprising Hayek's liberal societal order came down to self-interest in the broad sense; that is, the desire to pursue one's own objectives regardless of the particular motive. This is both individualistic and subjective; in Hayek's system of thought, it has a proactive meaning. As we see further on in the next chapter, Hayek harshly criticised the concepts of the 'emotions of a tribal society' and 'social justice' based on an altruism that assumes the specific values of a community. However, one must consider that this does not necessarily assume a dichotomy of self-interest versus altruism. What Hayek was concerned about were the various movements to apply to the limited and individual 'objectives' and values of specific people and groups in order to change the market society and its general rules. Hayek criticised all of these as various forms of 'constructivism', and at the same time, criticised the attitude of directly reducing social order to innate instinct regardless of self-interest or altruism. This is also the point where Hayek clearly departs from mainstream economics and it is the reason for his unique social philosophy.

The self-interest or individualism upon which Hayek based his liberal philosophy was not, for example, a pure Hobbesian egoism for self-preservation. Hayek rejected the method of building social order through contracts or designs that assumed atomistic entities. Hayek used terms such as 'self-love' and 'selfish interests', but these terms did not mean egotism in the narrow sense of concern, only with one's own immediate needs and wants. The 'self', for which people were supposed to care, did as a matter of course, include their family and friends' (Hayek 1949, p. 13). While the interests of people themselves are extended to a certain societal range out

of concerns that are based on one's own interests, Hayek emphasised that there are certain limits to that range. He continues:

Whether he is completely selfish or the most perfect altruist, the human needs for which he can effectively care are an almost negligible fraction of the needs of all members of society. The real question, therefore, is not whether man is, or ought to be, guided by selfish motives but whether we can allow him to be guided in his actions by those immediate consequences which he can know and care for or whether he ought to be made to do what seems appropriate to somebody else who is supposed to possess a fuller comprehension of the significance of these actions to society as a whole (Hayek 1949, p. 14).

It is certainly the case that Hayek's standpoint was a type of consequentialism, meaning that only a free society could generate and maintain maximum economic prosperity and population,<sup>3</sup> although he simultaneously criticised Benthamite utilitarianism that tries to design ideal legal norms and social systems on the basis of direct comparison and calculation of individual utility (Hayek 1976, p. 18). This is because of the importance of 'dispersed knowledge', which he emphasised throughout his life. Hayek did not assert that an economic society is directly established by self-interest. Rather, he criticised a direct reduction of the establishment of social systems to such elements, whether self-interest or altruism. Thus, the problem was 'not whether people are led by self-interested motivations or whether they should be led by such motivations' (Hayek 1949, p. 14). Assuming limited and 'dispersed' knowledge, we must focus on the process by which human behaviour shapes society based on such knowledge (i.e. spontaneous order). In other words, Hayek did not frame the problem of human nature as a matter of good or bad, or self-interest or altruism, but rather, he tried to move beyond such ideas. He stated that the issue was the importance of social order created by the cumulative actions of individuals, whether good or bad,

---

<sup>3</sup>Kukathas (1989) declared that Hayek's liberalism was 'end-independent' and was generally regarded as representative of 'negative liberalism'. Although in the sense that his objective was a higher level of self-realization with freedom as a means, Hayek was a type of Kantian and belonged to a kind of 'positive liberalism'. Further, he noted Hayek's failure to balance a Humean empiricism (or consequentialism in a broader sense) and Kantian deontology and argued that the latter should be excluded to focus on the former. On the other hand, while Kley (1994) partly acknowledges the assertions of Kukathas, he regards Hayek as a consequentialist who disguised himself in a mask of Kantian. Moreover, he counters that Hayek was actually reticent about the facts of value objectives that should be strived for, by simultaneously negatively assessing Hayek's eclectic attitude. A more detailed discussion of these interpretations cannot be given here—although we must acknowledge that Hayek placed a great distance between himself and simple Benthamist utilitarianism. See Yamanaka (2007, Chap. 2) for a more detailed discussion of this issue in Japanese.

I think that although these scholar's arguments are worthy of attention, they overemphasised the division between the Humean and Kantian characteristics in an attempt to reduce Hayek to one or the other, which risks losing sight of the true value of Hayek's uniqueness. Mounce (1999) argued that the characteristics of Hume's naturalism were his considerations of thoroughly worldly aspects without any concept of God and his denial of a metaphysical foundation since he denied a reduction to simple experientialism or empiricism, giving this the significance of naturalism that is differentiated from materialism. I interpret Mounce's assertion as fitting into the characteristics of Hayek's natural law theory as discussed in Sect. 4. See Taishido (2003, 2005) for the relevant discussion in Japanese. On the other hand, the argument on Chap. 4 (Itai's) of this book intends to go beyond the prevailing notion on Bentham's utilitarianism.

self-centred or altruistic, or rather a ‘spontaneous order’ that is ‘the result of human action, but not the execution of any human design (Hayek 1946, p.7)’. As we can see from the fact that he quotes Adam Ferguson, these concepts had been previously discovered by enlightened British thinkers, such as Mandeville, Hume, Smith, and Ferguson.<sup>4</sup>

## 2.2 *Economic Order as a Catallaxy*

The following type of economic society is what Hayek was advocating by means of the above type of human character:

It is a social system which does not depend for its functioning on our finding good men for running it, or on all men becoming better than they now are, but which makes use of men in all their given variety and complexity, sometimes good and sometimes bad, sometimes intelligent, and more often stupid (Emphasis mine: Hayek 1949, pp. 11–12).

Such a system is a particular type of a market order; Hayek called it a ‘catallaxy’ rather than an economy. The word ‘economy’ originates from the ancient Greek word for domestic affairs, *oikonomia*, as the definition of distributing resources based on certain values in closed organisations such as households or corporations. Hayek used the term ‘catallaxy’ to show the spontaneous exchange that emerges from the dispersed decisions of individuals whose existence is not based on a top–down structure. The term comes from the Greek word *katallattein*, which means ‘to exchange’, ‘to admit into the community’, and ‘to change from enemy to friend’ (Hayek 1976, pp. 108–109).

---

<sup>4</sup>Petsoulas (2001) is representative of research on the relationship between the Scottish Enlightenment and Hayek’s thought. However, she asserts that the theories of Mandeville, Hume and Smith did not directly precede Hayek’s spontaneous order, which was a pillar of ‘social evolution through a process of selection’. According to Petsoulas, while those three thinkers were biased toward the role of human intelligence in response to the explanations of the establishment of various rules and systems, Hayek took too lightly the role of human rationality in response to the appearance of explicit rules and systems through trial and error. However, these arguments by Petsoulas seem to underestimate the role of ‘immanent criticism’ in Hayek’s arguments (Hayek 1976, p. 118). As is the case with the important role of judges in the legislative process, which we see in Sect. 4, Hayek does not completely deny the role of rationality in social evolution. Hakkonsen’s interpretation of Hume given in footnote 13 has much in common with Hayek’s thought and is a more legitimate interpretation.

Some critics have pointed out a certain level of affinity between Hayek’s view of the individual and communitarian arguments. For example, Gissurarson (1987, pp. 193–194), Kukathas (1989, p. 222) and McCann Jr. (2002, pp. 355–356). Based on those arguments, Doi (2008) suggested the surprising closeness between the thought of Hayek and Alasdair MacIntyre, who is one of the representatives of communitarianism. The emphasis on sociality behind these agents is a distinguishing feature of Hayek’s liberalism.

This notion was a carryover from the economic calculation controversy of the 1930s,<sup>5</sup> where Hayek harshly criticised market socialists such as Oskar Lange, who attempted to make rational plans based on all relevant information. His critique was based on doubts regarding reductionism, which held that it was possible to ascertain and calculate all data as elements of society.

The important point about the catallaxy is that it reconciles different knowledge and different purposes that, whether the individuals be selfish or not, will greatly differ from one person to another. It is because in the catallaxy mean, while following their own interests, whether wholly egotistical or highly altruistic, will further the aims of many others, most of whom they will never know, that it is as an overall order so superior to any deliberate organisation: in the Great Society the different members benefit from each other's efforts not only in spite of but often even because of their several aims being different (Emphasis mine: Hayek 1976, p. 18).

A catallaxy has aspects of a highly competitive society, given that 'it becomes clear who will best fill the needs of others by their actions'. However, it is not a zero-sum game since ability, skill, and luck all cause the common pool to expand when the game is played by the rules. In this society, the overall pie is enlarged by people working to expand their own interests in the broadest sense and acknowledging increases or decreases in their reward as there are improved opportunities for more people. A catallaxy as a spontaneous order is not a static equilibrium, such as simple neoclassical economics, based on perfect information. It exhibits nothing more than a general 'tendency' toward an equilibrium. The essence of markets is that equilibria are constantly fluctuating, and as they do, people's 'expectations' change, leading to discovery and the growth of knowledge. Thus, Hayek emphasised the importance of 'dispersed knowledge' among individuals who are unable to centrally manage or ascertain that knowledge, along with the universal and intrinsic imperfections in human awareness. He argued that only market processes could aggregate 'dispersed knowledge' and put it to use in the development of society.

Private property rights are a required for exchanging this kind of knowledge in the market. The existence of these property rights allows people to adapt knowledge to their own purposes, and at the same time, the method of dispersal itself is a factor in forming individual preferences and individuality. However, private property rights are not based on natural rights as suggested by John Locke and modern libertarian philosophers such as Robert Nozick, but are only formed as customary rules or 'conventions'. To Hayek, it is important to separate these general rules of a catallaxy from natural 'emotions'. This is discussed again in Sect. 4.

In any case, this means that the market eliminates arbitrariness as the various intentions of individuals come to fruition as 'unintentional consequences'. Individuals rely primarily on price signals as they continually predict the consequences of their behaviour, although those consequences are not necessarily achieved. Gaps occur because of ever-present uncertainty and asymmetrical information; however, these

---

<sup>5</sup>Lavoie (1985) pointed out that the outcome over the economic calculation controversy in 1920–30 played a major role in the development of Hayek's thought and provided important implications for subsequent research on him.



gaps are major signposts guiding people to further predictions and behaviour.' Social justice', or a society that controls a closed 'economy' based on 'group self-interest' and attempts to capture special and vested interests, changes a catallaxy system that enables the free actions of individuals into a fight for pieces of the pie in a zero-sum game, as will be discussed in the next section. Therefore, Hayek argued for the need to properly reconcile the motivations of individuals based on the concepts of 'rule or law' and 'open society rules' to escape from such a system and change the economic game to a catallaxy.

### 3 The Danger of Group Self-interest Distorting Market Order

#### 3.1 A Criticism of 'the Mirage of Social Justice'

Hayek thought that direct, mutual assistance within a group, namely, altruism and reciprocity, is central to primitive family or village communities ('face-to-face or tribal societies'). Today, it is generally recognised that such emotions are a foundation for daily life in private spaces. Therefore, Hayek believed that antipathy toward a catallaxy, or the general rules of a market society are entrenched in modern society; we are continually nostalgic for the old world and idealise it such that we see the development of a market society as a form of corruption. However, Hayek criticised this conventional wisdom as follows:

The confusion prevailing in this respect is most clearly shown by the common identification of the terms 'altruistic' and 'moral', and the constant abuse of the former, especially by the sociobiologists, to describe any action which is unpleasant or harmful to the doer but beneficial to society. Ethics is not a matter of choice (Hayek 1979, p. 167).

Hayek here says two things. The first is the 'natural' and instinctive rebellion against general rules of market order and the second is the fact that the term altruism has two meanings: both (1) the prioritisation of the interest of others ahead of oneself and (2) the carrying out of those interests with disregard for one's will. In Hayek's understanding, in the sense of achieving the subjective objectives of the individual, there is no difference between self-interest and altruism. If it is based on the voluntary will of a person, then abandoning one's physiological desires and practising abstinence or altruistic behaviour can certainly be praiseworthy at the individual level, and it can also be important in the areas of religion and morality. At the very least, Hayek has no issue with participation in communities in which individual members have agreed to follow such norms. However, Hayek further pointed out that it is impossible to impose a single set of values on society overall. Despite this, there are sometimes attempts to build an 'ideal society' as opposed to individual wills, via 'command' and 'control'. However, for Hayek, such efforts suppress freedom and infringe on the value and dignity of individuals.

In Hayek's thought, socialism and fascism are typical examples of such systems, as is the United Kingdom's post-war welfare state created against the backdrop of the thought of Keynes and William Beveridge. Starting with *'The Constitution of Liberty'* (Hayek 1960, Chap. 7), Hayek placed awareness of the threat to freedom when democracy is not used properly as a social means, but rather for one's own purposes, at the forefront of his writing. Despite the self-evident existence of democratic governments comprising free political agents, and given the increasingly fractured interests and wills within legislative bodies that ended up undercutting freedom, Hayek took the appearance of welfare states quite gravely. His main points in criticising welfare states were warnings against the dangers of transforming democracy in such a manner.<sup>6</sup> Hayek thought that the appearances of 'altruism' or 'the emotions of tribal society' in specific forms were a result of movements demanding 'social justice', which, in turn, called for egalitarianism and distributive justice.<sup>7</sup> Market order as a catallaxy is, as David Hume stated, an 'artificial virtue', and because it is an impersonal order, it easily incurs antipathy. Hayek felt that the rebellion against market expansion in various forms was a result of people's natural feelings that were based on community reciprocity and mutual assistance. Thus, Hayek called movements based on 'social justice' (including fascism as well as socialism), as opposed to market order, 'atavism to tribe society', even if these movements had modern colourings on their surface.<sup>8</sup> However, because these are throwbacks, they are doomed to

---

<sup>6</sup>Hayek harshly criticised existing welfare states but did not deny the importance of general welfare policies, instead thinking of them as major elements in the 'Constitution of Liberty'. Hayek uses a great deal of space in the third part of the book of that same title (Hayek 1960) to discuss this topic in detail, including the minimum income guarantees and public pension system. For detailed arguments for this, see Taishido (2011). Further, unlike the logical positivists whom Hayek also harshly criticised, he did not deny the significance of moral or ethical concepts, or the concept of justice itself, and did not take the standpoint of moral relativism, as seen here:

The freedom to pursue his own aims is, however, at least as important for the complete altruist as for the most selfish. Altruism, to be a virtue, certainly does not presuppose that one has to follow another person's will. But it is true that much pretended altruism manifests itself in a desire to make others serve the ends which the 'altruist' regards as important. (Hayek 1973, p. 56)

<sup>7</sup>Hayek thought that distribution based on standards such as 'desert', 'merit', and 'need' through states with a centralised authority was arbitrary and could not be true justice (Hayek 1976, p. 84). The reason was that there were no clear limitations to demands based on 'social justice', as well as the fundamental difficulty of measuring those demands. Thus, in the end, such decisions on distribution could not help but be arbitrary. Moreover, such a system brings about human rankings based on a single, arbitrary standard and a strict meritocracy, placing pressure on freedom. Thus, Hayek criticized attempts by states and society to directly create justice as 'anthropomorphism' (Hayek 1976, p. 64).

<sup>8</sup>Carl Polanyi, an opponent of Hayek's, understood the emergence of socialism and fascism as resistance movements in response to 'self-regulating markets', although both he and Hayek were aware, to some extent, that market order did not arise naturally, but that it is artificial, and thus unstable. However, Polanyi was negative toward prolonging 'self-regulating markets' by means such as welfare policies, and aimed at creating an alternative social order, a kind of socialism, while Hayek viewed such attitudes as movements for 'social justice' or 'atavism' that would lead society

failure since single standards of value cannot be applied within an overall catallaxy. Individuals only act in accordance with their own limited conceptual framework and ‘dispersed knowledge’; a liberal society is otherwise impossible.

### 3.2 *The Evils Created by Organised ‘Group Self-interest’*

In Hayek’s arguments, the economic and the political realms stand in stark contrast to each other and the former is overwhelmingly superior. Based on Hayek’s epistemology (‘dispersed knowledge’) and his support of the market, it is easy to understand that he does not appreciate the role of centralised politics, though it is true that people commit errors even in the economic realm. Continual failures in various forms are part and parcel of the economic realm. However, on what basis is the economic realm considered superior? Hayek did not regard the market as a place where static equilibrium is always achieved. Rather, his perspective is that in a market society, individual mistakes and trial and error occur simply because there is no agent controlling the whole, creating the possibility of generating long-term interests. This decentralisation is the greatest reason for market advocacy. This requires general rules for the equal treatment of individuals and for mediation when interests conflict. The goal-seeking behaviours of individuals that are based on these rules ought to be praised, even though in a collective form, they arbitrarily modify general rules and attempt to attain special interests, thereby creating the risk of distorting the overall order. Hayek expressed this danger as follows:

While public indignation and in consequence also legislation has been directed almost entirely against the selfish actions of single monopolists, or of a few conspicuous enterprises acting in concert, what is chiefly threatens to destroy the market order is not the selfish action of individual firms but the selfishness of organized groups. These have gained their power largely through the assistance government has given them to suppress those manifestations of individual selfishness which would have kept their action in check (Underlines mine: Hayek 1979, p.89).

Here, Hayek is criticising organisations that put the greatest emphasis on the interests of their own closed group to the exclusion of all others. As a typical example of such ‘group self-interest’, Hayek notes the pressure of labour unions that seek to protect their own vested interests to the exclusion of non-union workers, although it is worth noting that many other groups, such as lobbyists, rent-seekers, and, especially, ‘corporate capitalism’ can also be included in this concept (Egashira 2007). Rather than acknowledging unlimited property rights, Hayek showed an awareness of the need to provide a proper legal framework that includes rights such as patents and trademarks, which are fundamental to liberal and market society (Hayek 1967).<sup>9</sup> On

---

to collapse. This marked a clear divergence in the views of the two. See Stanfield (1986) for more information on Polanyi.

<sup>9</sup>In addition, Hayek argues for the importance of financial policies, rule-based eminent domain for and public objectives, anti-monopoly measures, and inheritance taxes. With regard to monetary

corporate law, while Hayek respected individual shareholder rights, he was critical of extending them to non-natural entities (corporations) without limit and stated that the law must make a distinction between the two. Hayek feared that there is danger in a concentration of stock being favourable to large companies as it distorts market structures. In that sense, there is no need to extend the rights of individuals as natural persons to organised groups. That is why government can be obligated to protect individuals from such groups through proper frameworks. The pursuit of individual interest is tied to the interests of the overall society through proper laws and rules, and designing these rules to ensure that the system does not play into the hands of 'group self-interest' is of utmost importance.<sup>10</sup>

Hayek pointed out that large 'incorporated' firms and labour unions seek to exclude not only non-regular workers or small- and medium-sized businesses, but also general consumers, taxpayers, women, and the elderly, who are unorganised by vested rights and special interests, despite comprising a large portion of society. Further, Hayek criticised the demands of vested rights and special interest groups against the backdrop of 'anger', saying that they are easily tied to democratic systems and could acquire legitimacy.

What is not yet generally recognized is that the *real exploiters in our present society* are not egotistic capitalists or entrepreneurs, and in fact not separate individuals, but organizations which derive their power from the moral support of collective action and the feeling of group loyalty. It is the built-in bias of our existing institutions in favour of organized interests which gives these organizations an artificial preponderance over the market forces and which is the main cause of real injustice in our society and of distortion of its economic structure. More real injustice is probably done in the name of group loyalty than from any selfish individual motives (Emphasis mine: Hayek 1979, p. 96).

People's demands establish themselves as opinions of the majority that procedurally come to an agreement. There is fundamentally no problem with this if the power of the government is properly restrained by 'the rule of law', although in actuality, this can become the 'playball of group interests' (Hayek 1979, p.99), making it impossible to guess where the process will end and increasing the risk that it will lead to arbitrary changes in the general rules. To avoid collusion between such egoism and democracy, Hayek presented a unique parliamentary reform plan based on a two-party system that emphasised the active role of judges within the scope of the rule of

---

policies, Hayek later proposed a theory for liberalising the issuance of currency (the denationalisation of money); given the difficulty that central banks face in executing policies, even when the principle of monetary policies according to rules has been maintained.

<sup>10</sup>The foundation of such thinking by Hayek was the importance of the value judgement as 'freedom'. As long as that was his basis, it seems that there are not necessarily any inconsistencies between the theory of market design in contemporary economics and Hayek's thought. 'Design' strives to create a market framework and make it efficient by skilfully coordinating the parts of a system; in contrast, the 'construction' criticised by Hayek directly leads in the direction of a specific society. On the other hand, as was noted by Saint-Paul (2011), the recklessness of methods that lack value principles (such as the importance of freedom) and simply use the outcomes of behavioural economics or neuroeconomics for governance (such as libertarian paternalism) are forms of reductionism, and the accumulation of designs based on 'nudges' may result in a suppression of liberty.

law. We discuss this in the next section. Hayek criticised the collectivisation of self-interest in the political realm although he did not deny the value of democracy itself and fully acknowledged its significance. He did sound the alarm on parliamentary systems becoming divided against the backdrop of ever-present interest groups and becoming ‘bargaining’ democracies. In particular, Hayek pointed out that group self-interest was a result of the ‘emotions of tribal society’ deeply rooted within humans, as well as to the ‘non-domesticated barbarians in our midst’ (Hayek 1978a, p. 66).

Moreover, Hayek noted that governments have the tendency to protect group interests as a separate factor in preventing the function of a market economy. That is to say, ‘we shall see that the universal tendency of politics is to give preferential consideration to few strong and therefore conspicuous effects over the numerous small and therefore neglected ones’ (Hayek 1976, p. 122). Such policies easily become the vested rights of the means, with an example by Hayek being the full employment, financial, and progressive taxation policies of Keynesian economics. For this reason, Hayek continued to criticise Keynesian economics throughout his life.<sup>11</sup> Setting aside the intent of Keynes himself, Hayek challenged policies that prioritise organised interests because they do not apply to all those within a society and lead to special rights and the discriminatory treatment of others. He repeatedly emphasised his thinking on ‘equality under the law’ where rules are applied equally to all people and denounced a society where such equality is not established such that the strong have an advantage at the expense of the relatively weak. However, it is the realisation of the ‘rule of law’ that has made possible a society based on expanded exchanges that we have today. Hayek noted that such rules have built up an abundance of experience that is greater than what people living today notice are, and that this has been formed through a long process of trial and error. Contrary to the danger of the policies that protect ‘group self-interest’ in rendering the principles guiding the direction of human effort dysfunctional, the objectives of rules in a free society with equality under the law, as asserted by Hayek, are to prevent the arbitrary exercise of authority and to increase the likelihood of favourable opportunities for all.

---

<sup>11</sup>The dispute between Keynes and Hayek is generally well known—although research in the future will likely examine characteristics and the intellectual foundation shared between the two. As for their intellectual exchange, see Ebenstein, (2001) and Wapshott (2011). If we were to outline a few points of argument, the first one would be the position of these two men as being in resistance to a ‘collapse of liberty’ in the mid-war period, particularly as advocates for protecting Western civilization or a certain type of moral philosophers, instead of understanding the thoughts of Keynes and Hayek as a dichotomy. Second, both criticised the theories of equilibriums of mainstream neoclassical economics and were heretical economists who emphasised the imperfection of the knowledge of economic agents, as well as their sociality and inter-subjectivity when establishing that knowledge. On the premises of such assumption, the differences between Keynes and his unique macroeconomic assertions and Hayek and his opposition to them may also become apparent. Third, both laid emphasis on the customs for maintaining ‘civilisation’ or historical social order, and thus they both showed their strong devotion to British thinkers such as Hume and Burke. Of course, there are differences between the two. Hayek placed his hopes in the continuously accumulated customs and stability of market order based on ‘nomos’ as the law of liberty, while Keynes emphasised the uncertainty and baselessness (typically shown in the Keynesian beauty contest) of people’s beliefs upon which customs are based, as well as the significance of a certain degree of government intervention. I owe a lot to Mamiya (2006) on these points.

## 4 Hayek on the ‘Rule of Law’

### 4.1 *The Natural Hierarchical Structure of the Law*

In this section, we examine foundational thinking that supports spontaneous order as the ‘rules of an open society’, particularly focusing on the two characteristics of Hayek’s theory of the ‘rule of law’: natural law and republicanism. First, let us consider the characteristics of natural law. It is relatively well known that Hayek’s concept of rules is based on the dichotomy of ‘nomos’ and ‘thesis’. ‘Thesis’ are internal rules of specific organisations such as families, corporations, various types of associations, or government groups; in other words, ‘taxis’. While these delineate proactive, specific objectives and operating principles, nomos are rules for spontaneous order as a free society (cosmos) that incorporate thesis. They are passive in their establishment of property rights and the extent of private space, as well as in the banning of infringement of others’ actions and spaces. However, because they are passive rules, nomos are the source for the acknowledgement of free action by individuals and organisations (taxis). Further, Hayek emphasised the contrast between the two rules in response to thesis being based on the achievement of specific objectives (will), saying that nomos was ‘the view (opinion) about the desirability or undesirability of different forms of actions, or actions of certain kinds’ (Hayek 1976, p. 13). Opinions do not set forth specific objectives of people but are rather a higher concept of standards determining the propriety of objectives and an awareness of proper frameworks for markets and social activities. In addition, opinions are not clearly codified and are often not clearly explicated by those who have them; however, they exist as cumulatively created customs among people in society. Our actions and activities are defined by opinions and act as guides with regard to the propriety of those actions and activities.

However, various types of organisations (taxis) in the activities of society are of course required, and Hayek did not deny the significance of the existence of thesis as rules of organisations. He fully acknowledged the roles of thesis, but took issue with thesis that are nothing more than the rules of one organisation being applied and extended to the general society. Typical examples of this are the pre-modern patriarchal states and the societies controlled by modern fascism or socialism. The operation of states by thesis may have been possible in small populations or the polis of ancient Greece (i.e. face-to-face or tribal societies); however, Hayek believed that unlike these small populations, the operation of the state by thesis is impossible in ‘great societies’ with large populations with differing intents and objectives. In addition, Hayek’s criticism was not only of socialism or of fascism but was also deeply related to the method of legitimising liberal societies. Put differently, Hayek thought that regardless of the system, it was impossible to construct societal rules according to a single rationale and that the forceful application of rules under the guise of such a rationale or social justice could only be arbitrary.

In that sense, it is important to base rules on a ‘higher law’. In other words, the existence of thesis is accepted as they follow nomos, which are a guaranteed

legitimacy as they are built on opinions. It is significant to note that Hayek did not deny positive law as thesis through the dichotomy of *nomos* and thesis. Rather, Hayek carefully formulated the issue of identifying the foundation that positive law is built upon and the manner in which it is enacted. Hayek's concept of rules had a similar hierarchical structure yet also developed based on the body of thought propounded by Hume, Smith, and other enlightened Scottish thinkers, having an affinity to the non-theological modern theory of natural law.<sup>12</sup> Characteristic of his philosophy of law is, first, that *nomos*, or the opinions upon which *nomos* are premised, are based on the worldly, artificial actions of people, and at the same time, handed down over generations rather than being the will of a single person or generation. Thus, spontaneous order is 'the result of human action, but not the execution of any human design' (Adam Ferguson) and exists outside of the purposeful will of a single legislator; therefore, it is natural and not arbitrary.<sup>13</sup> Second, Hayek emphasised the importance of discovery in practical human actions, such that humans cannot grasp the entire picture directly, although this was also closely related to his market view, which emphasised the process of 'competition as a discovery procedure'.

In addition, while Hayek praised the Scottish enlightenment thinkers, he criticised the French social thinker Rousseau as a 'fake individualist' (Hayek 1949). In that context, Hayek kept in mind the contrast between Humean 'opinions' and Rousseauian 'will'. Hayek believed that not only 'volonté de tous', or total will, but also 'volonté generale', or general will—as long as it is oriented toward guiding society in a specific direction—will be the source of oppression and tyranny. It

---

<sup>12</sup>Much research already exists on natural law interpretations of Hayek's thought, for example, the literature noted below. While Covell (1992) rejected transcendental natural law theory (Platonism and modern rationality), he regarded the jurisprudence of Hayek as a kind of natural law that pre-exists positive law as 'formal, procedural principles'. However, I think that Hayek's natural law goes beyond the 'formal, procedural principles' espoused by Covell. van Dun's (1994) work is an interpretation that emphasises the similarities with Humean natural law, and Roos's (1994) work is an interpretation that emphasizes the similarities with Kantian natural law. Cliteur (2000) defined the characteristics of natural law as (1) the existence of legal restraints; (2) adherence of positive law to a 'higher law'; and (3) naturalism. Hayek's theory of law is thought to be a developmental natural law theory that meets all of these definitions. See Taishido (2005) for the discussion in Japanese. Angner (2007) also viewed the basis of Hayek's theory of spontaneous order as the superiority of natural law over artificial law.

<sup>13</sup>Although the relationship between the thoughts of Hume and Hayek cannot be treated enough here,— See Taishido (2003, 2005) for discussion in Japanese—I think Haakonssen's perspective has much in common with Hayek. Haakonssen (1981) says the following about Hume's theory of justice.

the particular boldness in Hume is that he uses it (the idea that many social phenomena are the unintended consequences of human actions: added by the quoter) in accounting for one of the traditionally most central, and in a way most 'sacred', elements in social life, namely fundamental law itself, our very 'sense of justice'. It is of the most important parts of his philosophical justification for replacing traditional natural law with a secular and empirical conception of fundamental law...it does, in common view traditional natural law theories, find roots of justice beyond any rational human deliberations, and far beyond our present society (Haakonssen 1981, p. 21).



must be noted that, in Rousseau's arguments as well, total will that is the simple sum and aggregate of particular will as the individual's interest cannot become a proxy for the common interest of society, and is, therefore, nothing more than an arbitrary existence. This is different from the 'general will' that is a fair guide for conducting society, and which society holds overall, separate from the interests of citizens that constitute a community (state) established on the basis of a social contract. Removing arbitrariness and establishing fair general rules is a point that is common to both Hayek and Rousseau. However, what decidedly differentiates the two are opinion and will. The opinion espoused by Hayek, or rules as *nomos* based on opinions, are not commands to people for direct action but help in guiding free activity. Conversely, Hayek thought that no matter how commonly held a 'general will' is, governance that is based on it cannot become a foundation for a free society, given that the absolute submission to national sovereignty it required. Thus, it is very interesting that in the end, Hayek protested that the shift from the divine right of kings to popular sovereignty was not historical progress that necessarily expanded freedom, but rather strengthened the subordination of people to a sovereign state. In consequence, he suggested that we abandon the concept of 'sovereignty' because it has an aspect of absolute subordination to command (Hayek 1979, p. 124).

#### ***4.2 Hayek's Ideas on Parliamentary Reform***

Hayek proposed a unique theory of legal institution for discovering or establishing rules as *nomos*, first pointing out the importance of the role of judges. Naturally, judges issue decisions when the behaviour of individuals and organisations come into conflict and the situation worsens into a dispute—although they do not consider all of the reasons for the dispute. Their standard is existing opinions, which are the bases of *nomos*, and are shown as accumulated precedents. What judges do is to modify existing systems little by little, not to construct systems in their entirety. Hayek emphasised that when doing so, judges are required to make a 'deliberate effort' (Hayek 1973, p. 100), and his thinking, which vigorously denied and described the arbitrary laws of government officials and their application as constructivism, has very interesting connections to this effort. However, judges deliberate to remove arbitrariness, and their discretion is always predicated on existing opinions and *nomos* and is regulated by them. The characteristics of Hayek's theory of the rule of law are clear in the following:

The rule of law is therefore not a rule of the law, but a rule concerning what the law ought to be, a meta-legal doctrine or a political ideal. It will be effective only in so far as the legislator feels bound by it. In a democracy, this means that it will not prevail unless it forms part of the moral tradition of the community, a common ideal shared and unquestioningly accepted by the majority. (Hayek 1960, p. 206)

It is impossible for democracy to become established without the rule of law; it always requires a higher concept to control it. Hayek also emphasises appropriate



legislative reform as well as the role of the judges, although he constantly warned of the corruption of democracies, as they become ‘bargaining’ democracies through interest groups (Hayek 1979). As has been mentioned, Hayek feared that governance based on thesis or the will of such groups would come about. He proposed the following bicameral model to prevent this corruption and to solve the dilemma of maintaining democracy and the division of power.

First, Hayek named the lower house chosen in a regular election of all male and female voters, just as we normally conceive, the ‘governmental assembly’. He had given up the legislation of *nomos* enacted through such a normal assembly and asserted that it could not avoid becoming a place of continual rent-seeking and other conflicts of interest. Thus, he limited it to the enactment of thesis, namely the specific selection of public services by the government in response to annual budgets and the setting of rules for enforcement. On the other hand, Hayek named the upper house the ‘legislative assembly’ and assigned to it the role of setting standards—taxation, labour, production, construction and food safety, etc.—and general rules, or *nomos*, for the market society, such as corporate law. The standards for selecting members are unique, as they are selected for 15-year terms from among men and women of age 45–60 who have achieved notoriety in society and are ‘people who have already proven themselves in the ordinary business of life’. Such extremely long terms are in place to ensure that the short-term interests of the electorate do not influence the judgements of assembly members as they strive for re-election, with one-fifteenth of seats up for re-election each year. In addition, members who are selected once for an administrative assembly do not have eligibility for election. Hayek intended a radical decentralisation of power through such a system.

Moreover, Hayek proposed using the term ‘demarchy’ instead of democracy to express a democracy properly limited by the rule of law. The word ‘democracy’ is a combination of ‘*demos*’ (the masses) and ‘*kratos*’ (power) and was viewed by Hayek as denoting unlimited power via a will that was oriented toward specific objectives. On the other hand, ‘demarchy’ combines ‘*demos*’ with ‘*archein*’ (control) and takes on the meaning of control via established laws that are well known to the people and not by arbitrary directives (Hayek 1967, p. 96).

While Hayek’s plan for reform is, of itself, fascinating, it is unclear as to the manner in which the eligibility of individuals is determined, and many would doubt whether it is appropriate to give major authority to assembly members for a long 15-year term. Hayek does not present a detailed and specific proposal, and doubts remain as to whether anything in his proposal is actually adaptable. Feasibility aside, however, the venture to create an authoritative upper house based on a completely different system to restrain the powers of a lower house has strong shades of republicanism. As previously discussed, Hayek denied that political ‘will’ unites the overall society and was rather negative toward political participation having any specific significance. His intention was to separate ‘*res publica* (public affair)’ from party disputes and conflicts of interest and to create a discussion forum that would ensure impartiality—not swayed by such disputes.

We can summarise the above characteristics of Hayek’s rule of law in three points: resistance to corruption, removal of arbitrariness, and the creation of a legislative

organ that achieves them. In that sense, we can call Hayek's standpoint one of 'negative republicanism', being related to Philip Pettit's republicanism theory, as discussed in the next section.<sup>14</sup>

### 4.3 Hayek and Two Forms of Republicanism

It is well known that current discussions around republicanism are mainly of two types. The first is classical republicanism, which is based on the civic humanism presented by Pocock (1975). In addition to having the proactive participation of citizens in the public space or governments as the foundation of its existence and preventing the corruption of tyrants or fallen oligarchy and democracy, it primarily identified—in historical thought—the aim to achieve the perfection of human character. Hayek can certainly not be thought easily compatible with this viewpoint.<sup>15</sup>

However, the second tradition of republicanism considers actual systems design and the governance theory to prevent the corruption of social systems. This tradition is not necessarily in conflict with the idealism noted in the Pocock's discussion. The theory of the ideal state of Aristotle and Cicero, who were advocates of classical republicanism, is simultaneously a theory of mixed bodies to halt the corruption of power; in other words, it has a strong balance of power that combines elements of monarchy, aristocracy, and democracy. In the traditions of republicanism propounded by Skinner (1998) and Pettit (1997), emphasis was laid on institutional arguments of law and governance. Skinner's primary interest lay in a historical analysis of thought, while Pettit developed more practical arguments. Like Skinner, Pettit argued for 'freedom as non-interference' as a characteristic of negative liberalism, presenting the more sophisticated idea of 'freedom as non-domination' as a principle of resistance against arbitrary power that goes beyond 'freedom as non-interference'. Moreover, he suggested the possibility of mediating the dispute between liberalism (libertarians and liberals) and communitarians.

Pettit argued for three conditions of constitutionalism when creating the ideal government for preventing arbitrary interference or power. These include (1) an 'empire of law' in the sense of a fully realised principle of the rule of law and not the

---

<sup>14</sup>Honohan (2002, p. 8) classified republicanism into two categories, calling the proactive type that focuses on participation in the government space 'strong republicanism' and calling the type that places importance on individual liberty and the theory of institution as a foundation for that liberty 'instrumental republicanism'.

<sup>15</sup>Nevertheless, Hayek sent the following praise for Pocock's *The Ancient Constitution and the Feudal Law* (1957) for his consideration of English common law and its incorporation in his own arguments.

I would like to make amends here for inadvertently not referring to this excellent book in *The Constitution of Liberty*, 1960, for the final revision of which I had much profited from Mr. Pocock's work. (Hayek 1978b, p. 256)

rule of the people, (2) a separation of powers through the creation of different organs of power, and (3) an anti-majoritarianism that prevents despotism by majority rule and minority oppression, without reducing decision-making to simple headcount. In addition to rejecting legislation that is based on specific party interests, this type of institutional framework ensures freedom (Pettit 1997, pp. 171–183). The liberal concepts and system of ideas are similar to Hayek's position and share a certain affinity in that they remove arbitrariness.

In recent research, Kacelenbogen (2011) identified the closeness of Pettit's theory of republicanism to Hayek's thought. First, both have an awareness of the grave risk of the arbitrariness of interest groups to governance. Second, both point out the need for a limited role of government in operating market systems properly through legal institution. Third, both emphasise the role of customs when laws are formed to ensure liberty. Kacelenbogen notes that both men stress an evolution of cumulative knowledge that is not constructivism or the same 'modest epistemological position' (p. 463). However, as Kacelenbogen also points out, the two have major gaps between their standpoints. These men were differentiated by the 'contestatory democracy' emphasised by Pettit (1997, p. 185). He asserted that it is not only the establishment of a uniform authority through general elections, but rather the formation of a specific rights and institution for continually registering dissent against that authority that is required to remove arbitrariness and achieve the fair rule of law. Such debates are the conditions for a 'deliberate republic' with no arbitrary control (Pettit 1997, p. 187).

However, Hayek did not emphasise the efficacy of a direct debate process. Debates themselves fill a role; to a certain extent, they foster the opinions the society in question seeks. However, such decision-making should not be employed in overall policymaking. Rather, the work of abstraction and of making decisions consistent with long-term customs is more important, hence the emphasis on the role of judges and limits on membership in an upper house, the legislative assembly, as described above. In that sense, Hayek, though not ignoring the political process, prioritised limiting its influence and promoted adherence to a legislative process. Hayek's effort occurs not in direct participation in debates or in quickly reflecting it in policies, but in the question of how to indirectly handle such participation.<sup>16</sup> Against this

---

<sup>16</sup>On this point, it seems that there are major differences between the liberalisms of Hayek and Frank Knight. There are two main points, based on the differences in their views of 'discussion' and 'social justice'. To Knight, the concepts of freedom and human rights are more or less products of historical coincidence, and because of this, public 'discussions' play important roles when setting rules that advocate for the ideals of a free society. In addition, gaining consensus requires discussions on 'social justice', where there are individual and specific views on issues such as poverty measures and taxation, etc. Based on this perspective, Knight criticised that Hayek rejects 'social justice' and, for the most part, does not rely on the political process of discussions, instead entrusts the fate of the growth of liberal society to the extremely unstable process of natural selection. However, Hayek thought, as noted in this chapter, that distribution under the concepts of 'social justice' is arbitrary and can never be true justice. In a 'great society' made up of countless people with diverse values, the issue of distributive justice as to who gets how much cannot be solved through 'social justice' or the 'emotions of a tribal society', which are only the norms for certain small groups and organisations. A framework of abstract rules separated from 'social justice' is required to create a

backdrop, Hayek's unique epistemology held knowledge to be something dispersed among individuals, impossible to ascertain in totality and manifesting itself only as 'the unintended consequences' of market activities. However, although different in that idea, Hayek and Pettit's perspectives could be said to have commonalities: first, the theory of governance for achieving the rule of law that is not reduced to a simple majority decision; second, the removal of arbitrary authority; and third, the emphasis on the evolution of cumulative knowledge.<sup>17</sup>

In our discussions up to this point, we have considered the republican moments in Hayek's thought. These elements, which are characteristic of Hayek's theory of the rule of law, largely overlap with the 'instrumental', or negative, republicanism relied upon by Pettit, or to a greater extent, John Rawls. However, it is interesting that Hayek in the end argued for 'the dethronement of politics' (Hayek 1979, p. 147) in Chap. 18 of *Law, Legislation and Liberty*, along with disposing of the concepts of sovereignty, and proposed a unique theory of institution for proper legislation that was based on the separation and suppression of control and the removal of arbitrariness. Hayek felt the risk in the continual bloating of government structures and their influence expanding into various areas. He thought that overall control through specific value standards via the ideas of 'social justice' could grow within a democracy, with the danger of eroding and collapsing it. For 'the last battle against arbitrary power' (Hayek 1979, p. 152), there was a need for a rule of law that was based on the long-term principles and systems intended by Hayek.

## 5 Conclusion

As we have seen, Hayek stood on a horizon that relativised the conflict between self-interest and altruism, revolving around the importance of individuals' pursuit of goals against the backdrop of decentralised and dispersed knowledge. Characteristic of Hayek's thought was, first, his attempt to talk about new areas that transcend the dichotomy of self-interest and altruism. Hayek used dichotomies in his own arguments: *nomos* and *thesis*; *cosmos* and *taxis*; rules and commands; opinions and

---

multidimensional society that is not controlled by specific value standards. See Kern (1985) and Emmett (2007) for more information on Knight's criticism of Hayek.

<sup>17</sup>Pettit himself regarded Hayek not as a representative of 'freedom as non-domination' theory, but as that of 'freedom as non-interference', and attempted to differentiate it from his own argument (Pettit 1997, p. 89). Through this differentiation, Pettit emphasised the differences between the concepts of negative liberalism in the general sense—particularly those of Hobbes and Bentham—and his own republican liberty, bringing into relief the uniqueness of the latter. However, I think that Pettit not paying attention to Hayek's argument as differing from negative liberalism in a narrow sense is problematic. Even Isaiah Berlin, the father of the dichotomy of negative liberalism and positive liberalism did not make a goal of value relativism in simply praising the former and treating the preferences of all individuals equally (Berlin 1969). Rather, he acknowledged the incommensurability of individual values and emphasised pluralism, the goal of which is not the relativism of indifference to the values and ideals of others, but the mutual recognition and tolerance of them. Berlin's negative liberalism assumes this sort of pluralism.

will; freedom and construction; and the great society and tribal (or face-to-face) society. The list goes on. Such dichotomies are not used to simply praise one and condemn the other, but have dialectical shades, as Sciabarra (1995) argued. In other words, Hayek continually set two concepts in opposition to illuminate something beyond them. Characteristic of such dialectical thinking by Hayek was that although social order is based on the intention of each person in each aspect, it results in ‘unintended consequences’ of such behavior. Such attitudes of Hayek recall Hume in his political writings, who explained reconciliation in party battles through the two-front criticism of Tories and Whigs, and who stood on his own philosophical viewpoints by rejecting both simple scepticism and metaphysics. Their thoughts were unique in that regard.

Second, Hayek idealised a society in which individuals’ motives, whatever their form, lead to positive results through their optimization behaviors, and emphasised the importance of general rules for achieving this ideal. Nevertheless, Hayek roundly criticised the danger of distorting the generality of laws and social order through ‘group self-interest’ that attempts to escape from the rules using special interests and the assertions of ‘social justice’ in an attempt to directly construct an altruism-based ideal society. Therefore, a proper legal system design is needed to establish market order as a catallaxy in order to properly reconcile various individual interests.<sup>18</sup> Put differently, self-interest that is not controlled by a proper system carries the risk of collapsing society through ‘group self-interest’. Thus, Hayek sought rules for a ‘great society’ created from various values that differed from single one in a ‘face-to-face society’. Moreover, the purpose of these kinds of rules was to establish a market framework that also included criticism of the unrestricted expansion of corporate capitalism and the necessity of appropriate corporate law and anti-monopoly law (this includes general welfare policies as the ‘constitution of liberty’ although it is beyond the scope of this chapter: see Taishido 2011).

Third, Hayek’s theory of the ‘rule of law’ was supported by characteristics of both natural law and republicanism with regard to the problem of what laws are proper for a liberal society.<sup>19</sup> As was discussed in Sect. 4, while Hayek placed extreme importance on the legislative process, including system design, for creating the conditions for

---

<sup>18</sup>Although these arguments of Hayek, in terms of whether the establishment of a liberal society is legitimised through the ‘evolutionary’ process, which he relied upon in later years, have often been criticised (e.g. Vanberg, 1986), his thought on a theory of knowledge, criticism of constructivism, and advocacy of progressive social development extended beyond a political standpoint had a certain impact on various parties, including leftists in the modern British Labour Party (Griffiths 2014). In addition, Sciabarra (1995) saw commonalities between Marx’s criticism of utopian socialism and Hayek’s criticism of constructivism because both criticised that utopianism has internalised the significance of an anti-historical abstraction of human possibility. Further, Sciabarra interpreted that both were ‘radical practitioners’ and that the positions of both, ostensibly on opposite ends of the spectrum, as being complementary.

<sup>19</sup>In the study of the history of ideas, the consideration of the relationship between the natural law tradition and the republican tradition is controversial. While Pocock (1981) distinguished between them and insisted on the autonomy of the latter (especially civic humanism), Skinner (1998) was critical of this and considered the neo-Roman ‘Liberty before Liberalism’ as the fusion of these two traditions. Similarly, Robertson (1985) emphasised the unity and inseparability of these traditions.

establishing markets, he tried to eliminate influence of the political process on the legislative process as much as possible. In this as well, Hayek did not necessarily deny the possibility of ‘market failures’ or markets falling into a disequilibrium as he thought that replacing the outcomes of market processes with political processes was impossible and that doing so would instead increase uncertainty. Further, from the perspective of ‘higher law’, the characteristic of Hayek’s theory of law as a natural jurisprudence restricts positive laws, filling the role of providing some basis to them. In addition, his own republican characteristics suppress the threat to liberal society posed by ‘group self-interest’ morphing into arbitrary power. These characteristics give Hayek a special place among liberal economists and social philosophers.

## References

- Angner, E. 2007. *Hayek and Natural law*. Routledge.
- Berlin, I. 1969. *Four Essays on Liberty*. Oxford University Press.
- Caldwell, B. 1988. Hayek’s Transformation. *History of Political Economy* 20(4): 513–541.
- Caldwell, B. 2004. *Hayek’s Challenge: An intellectual biography of F.A. Hayek*. University of Chicago Press.
- Covell, C. 1992. *The Defense of Natural Law: A study of the ideas of law and justice in the writings of Lon L. Fuller, Michael Oakeshot, F.A. Hayek, Ronald Dworkin, and John Finnis*. Macmillan.
- Cliteur, P.B. 2000. ‘Spontaneous Order, Natural Law, and Legal Positivism in the Work of F.A. Hayek. In *Hayek Revisited*, ed. Boudewijn Bouckaert, Annette Godart-van der Kroon, Edward Elgar: 14–31.
- Doi, T. 2008. ‘Hayek’s Theory of “Tradition of Rules of Conduct”’: Comparison with MacIntyre’s Theory of “Tradition of Intellectual Enquiry”’. *Chukyo Hogaku* 42(3 and 4): 173–184.
- Van Dun, F. 1994. Hayek and Natural Law: The Humean connection. In *Hayek, co-ordination and evolution: His legacy in philosophy, politics, economics and the history of ideas*, ed. by Birner, Jack, and Rudy van Zijp. Routledge: 269–286.
- Ebenstein, A. 2001. *Friedrich Hayek: A biography*. St. Martin’s Press.
- Egashira, S. 2007. Hayek on Corporate Capitalism. In *What is the market society? Vision and design*, ed. by Toshiaki Hirai. Sophia University Press (in Japanese): 116–139.
- Emmett, R.B. 2007. Knight’s challenge (to Hayek): Spontaneous order is not enough for governing a liberal society. In *Liberalism, conservatism, and Hayek’s idea of spontaneous order*, ed. Louis, Hunt and Peter McNamara. Palgrave Macmillan: 67–86.
- Fleetwood, S. 1995. *Hayek’s political economy: The socio-economics of order*. Routledge.
- Gissurarson, H.H. 1987. *Hayek’s conservative liberalism*, Garland Publishing.
- Griffiths, S. 2014. *Engaging Enemies: Hayek and the left*. Rowman & Littlefield.
- Hayek, F.A. 1944. *The road to serfdom: The definitive edition*: Edited with a Foreword and Introduction by Bruce Caldwell, University of Chicago Press.
- Hayek, F.A. 1947. ‘Free’ Enterprise and Competitive Order. In *Individualism and economic order*. Routledge & Kegan Paul: 107–118.
- Hayek, F. A. 1949. ‘Individualism: True and false’, in *Individualism and Economic Order*. Routledge & Kegan Paul.
- Hayek, F.A. 1960. *The constitution of liberty*. University of Chicago Press.
- Hayek, F.A. 1961. The non sequitur of the ‘Dependence Effect’. *Southern Economic Journal* 27(4): 346–348.
- Hayek, F.A. 1967. The corporation in a democratic society. In *Studies in Philosophy, Politics and Economics*. University of Chicago Press: 300–312.
- Hayek, F.A. 1973. *Law legislation and liberty: Rules and order*, vol. 1. University of Chicago Press.

- Hayek, F.A. 1976. *Law legislation and liberty: The mirage of social justice*, vol. 2. University of Chicago Press.
- Hayek, F.A. 1978a. The atavism of social justice. In *New studies in philosophy, politics, economics and the history of ideas*. University of Chicago Press: 57–70.
- Hayek, F.A. 1978b. Dr Bernard Mandeville. In *New Studies in Philosophy, Politics, Economics and the History of Ideas*. University of Chicago Press: 249–266.
- Hayek, F.A. 1979. *Law legislation and liberty: The political order of a free people*, vol. 3. University of Chicago Press.
- Honohan, I. 2002. *Civic republicanism*. London: Routledge.
- Kacenelebogen, E. 2011. Epistemological modesty within contemporary political thought: A link between Hayek's neoliberalism and Pettit's republicanism. *European Journal of Political Theory* 8 (4): 449–471.
- Kern, W.S. 1985. On the market as a game: Hayek vs. Knight. *R. H. E. T. M.*3: 51–59.
- Kley, R. 1994. *Hayek's social and political thought*. Oxford University Press.
- Knight, F.H. 1947. *Freedom and reform: Essays in economics and social philosophy*, New ed. Indianapolis: Liberty Press.
- Kukathas, C. 1989. *Hayek and modern liberalism*. Clarendon Press.
- Lavoie, D.C. 1985. *Rivalry and central planning: A re-examination of the debate over economic calculation under socialism*. Cambridge University Press.
- Mamiya, Y. 2006. *Keynes and Hayek: Transformation of 'Liberty'*. Chikuma Shobo Publishing (in Japanese).
- McCann Jr., C.R. 2002. F. A. Hayek: The liberal as communitarian. *The Review of Austrian Economics* 15: 5–34.
- Mounce, H.O. 1999. *Hume's naturalism*. Routledge.
- Pettit, P. 1997. *Republicanism: A theory of freedom and government*. Oxford University Press.
- Petsoulas, C. 2001. *Hayek's liberalism and its origins: His idea of Spontaneous order and the scottish enlightenment*. London: Routledge.
- Pocock, J.G.A. 1975. *The Machiavellian Moment: Florentine political thought and the atlantic republican tradition*. Princeton University Press.
- Robertson, J. 1985. *The Scottish enlightenment and the militia issue*. John Donald.
- Roos, N.H.M. 1994. Hayek's Kantian heritage and natural law. In *Hayek, co-ordination and evolution: His legacy in philosophy, politics, economics and the history of ideas*, edited by Jack Birner and Rudy van Zijp, Routledge.
- Saint-Paul, G. 2011. *The tyranny of utility: Behavioral social science and the rise of paternalism*. Princeton University Press.
- Sandel, M. 1984. The Procedural Republic and the Unencumbered Self. *Political Theory* 12 (1): 81–96.
- Sciabarra, C.M. 1995. *Marx, Hayek and Utopia*. New York: State University of New York Press.
- Skinner, Q. 1998. *Liberty before Liberalism*. Cambridge: Cambridge University Press.
- Stanfield, J.R. 1986. *The economic thought of Karl Polanyi*. New York: St. Martin's Press.
- Taishido, M. 2003. Hayek and Hume, Smith: Around the formation process of the social order. *Annals of the Society for the History of Economic Thought* 43: 52–67 (in Japanese).
- Taishido, M. 2005. Hayek on the notion of nature and natural law. *Keizai-Ronsō (The Economic Review) Kyoto University Economic Society* 175 (5–6): 486–503 (in Japanese).
- Taishido, M. 2011. Hayek's criticism of welfare states and his alternative vision: What conditions are required for ordering the free market? In *Poverty and welfare in the history of economic thought (Keizaishiso nonakano Hinkon, Fukushi)* ed. Atsushi Komine. Minerva Shobo Publishing (in Japanese): 193–230.
- Vanberg, V.J. 1986. Spontaneous market order and social rules: A critical examination of F. A. Hayek's Theory of Cultural Evolution. *Economics and Philosophy* 2 (1): 75–100.
- Wapshott, N. 2011. *Keynes Hayek: The clash that defined modern economics*. New York: W. W. Norton.

Yamanaka, M. 2007. *Hayek's political thought (Haieku no seiji shiso)*. Keiso Shobo Publishing (in Japanese).



# **Contemporary Topics in Self-Interest and Economics**

# Creating an Algorithm Based on the Theory of Moral Sentiments



Susumu Egashira

**Abstract** This chapter examines Adam Smith’s argument in *The Theory of Moral Sentiments* (TMS) based on a cognitive model. It considers its algorithm and builds an agent-based model. The TMS contains rich suggestions because it argues on how morality has emerged in a society in which selfish individuals interact. Many previous studies have evaluated it highly in the history of economic thought. However, few studies precisely discuss the logic of transitioning from being “sympathetic to an impartial spectator and the relationship between “an impartial spectator” and “morality”. The chapter aims to reevaluate Smith’s argument based on modern cognitive science and an agent-based simulation. Although the theory’s basis differs from the modern artificial intelligence, it examines the emergence of a socially favorable order in the interactions of individuals as does the modern social science. This chapter discusses the potential of the TMS in our society by modeling it in accordance with modern cognitive science.

**Keywords** Adam Smith · The theory of moral sentiments · Agent-based model · Algorithm

## 1 Introduction

This chapter examines the concepts of sympathy and an impartial spectator as stated in *The Theory of Moral Sentiments* (TMS, 1759) and considers an algorithm developed for it in an agent-based computer simulation model. Although the TMS was published long before *The Wealth of Nations* (WN, 1776), it discusses how morality has emerged in a society which consists of selfish individuals as assumed in the WN. Smith asserts that individuals successfully developed an impartial spectator that is based on the well-known concept of sympathy, which is an aspect of human nature.

Despite its significance in the history of economics, many previous studies on Smith have discussed the logical process of the TMS in his argument. However,

---

S. Egashira (✉)  
Otaru University of Commerce, Otaru, Japan  
e-mail: [egashira@res.otaru-uc.ac.jp](mailto:egashira@res.otaru-uc.ac.jp)

the process of the emergence of morality from the two aspects of human nature, selfishness and sympathy, is not self-evident. It is possible that economic development based on both self-interest and morality is not feasible despite all individuals having an “impartial spectator.” First, the society observed by Smith was that before The Industrial Revolution where most workers were poor and uneducated. Smith’s concern was to explain emergence of morality in a way scientific way. Smith did not regard emergence of morality as self-evident, and social order did not emerge spontaneously through human nature. Studies on the history of economic thought deem the logical consistency of classical material unimportant. However, our society has become much more complicated, and the interactions between humans and other agents are incomparable with Smith’s era. Furthermore, agents with artificial intelligence will soon begin interacting with humans and other AI. It is urgent to consider what kind of morality will emerge in the near future.

Needless to say, the argument in the TMS does not apply to modern artificial intelligence, which are based on big data and the Bayesian subjective probability. Moreover, the concept of human cognition in the TMS also differs with the modern one. However, despite acknowledging these differences, the TMS is valuable enough to be discussed because Smith was the first philosopher to convincingly explain the mechanism underlying the emergence of social rules. An argument that explains the emergence of morality with the fewest assumptions about human nature, with self-interest and sympathy as the premise for the essential discontinuity between individuals who are not able to immediately read each other’s minds, is compatible with an agent-based model. In this chapter, we converted the argument in the TMS into an algorithm that can be coded as a computer program. This procedure will clarify the shortcomings of Smith’s original argument. The concept on the impartial spectator is particularly important yet ambiguous. There are few explanations on how objective observations are processed in individual cognition. However, Smith did not provide a clear logical basis for neither the mechanism of sympathy, in which one shares the same emotion as another, nor the emergence of an impartial spectator.

Although the method employed in this chapter differs from Smith’s traditional studies, the philosophical reevaluation is consistent with Smith’s. This is a starting point for considering the emergence of morality in our society.

## 2 The Basic Cognitive Model in the TMS

This chapter is not the first to reevaluate Smith’s argument in light of modern cognitive science (Ashraf 2005). However, we constructed our model step-by-step according to each argument stated in the TMS because the translation into an agent-based model is unprecedented.

This form of interpretation is not standard in studies on Smith. However, Smith explicitly considered logic as important in his lecture on rhetoric (Smith 1983). Moreover, his argument is more logical than that of the enlightenment thinkers of his

age. Therefore, it is not entirely wrong if we adjust the argument stated in the TMS based on modern interest.

First, we reconfirmed Smith's theme setting and the conception of sympathy. Humans have no immediate experience of what their counterparts feel. Therefore, it is difficult to perceive the manner in which they are affected, except by conceiving what we would feel in a similar situation (Smith 1759[1976], 48). The only way through which a person can feel another person's emotion is by imagining what and how they would feel had they been in the same situation. Smith suggested that there is a cognitive break that separates individuals and that there is no way to objectively confirm that an image of one person's situation is similar to another person's. This means that Smith denied the possibility of co-experience between people and asserted that we can only perceive each other's subjective view through imagination.

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it. Of this kind is pity or compassion, the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner. (ibid., 48)

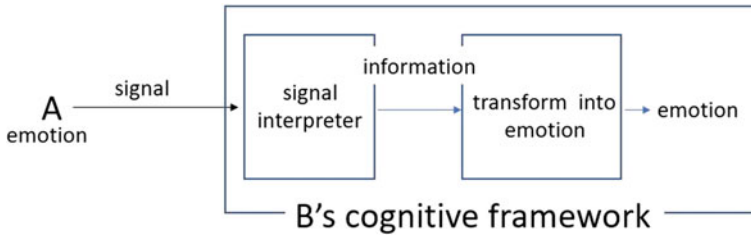
The first paragraph of the TMS contains more important factors than the concept of the isolation of individuals. First, this paragraph clarifies that what Smith aimed to explain in the TMS is related to human cognition. Cognition is a series of actions that involve the observation of a person and their circumstances, interpretation of the information, and conversion of this result into an emotion. To clarify our argument, a series of stimuli perceived by a sense organ is called a "signal." The resultant image of interpreting the signal is called information, while the psychological state elicited by information is defined as an "emotion." This process is described as a stream in which the first transformation interprets the signal and changes it into information, while the second transformation converts this information into an emotion.

It is not clear what kind of process should be assumed in these transformations, for example, statistical processing such as the Hebbian learning rule, because Smith did not explicitly describe it (Harry, Foundalis, and Martínez 2007). Therefore, we did not define these transformations here. It is not too much to say that one third of the first half of the TMS focuses on the second transformation, and arguments on the first transformation suggest an assumption of commonality between individuals.

Figure 1 represents this process. In the case of a mother watching a baby cry, the cognitive process is divided into an objective stimulus (a crying baby) > an observer (mother) > the first transformation (crying => feeling pain) - > the second transformation (negative emotion felt by the mother). The TMS hardly mentions the diversity of the interpretation of signals expressed by the first transformation.

It is important to note that Smith's concept of sympathy, its degree and quality, is dependent not only on an objective stimulus but also on its circumstances. The degree and quality of sympathy is dependent on these factors.

Even our sympathy with the grief or joy of another, before we are informed of the cause of either, always imperfect. General lamentations, which express nothing but the anguish of the sufferer, create rather a curiosity to enquire into his situation, along with some disposition to sympathize with him, than any actual sympathy that is very sensible (ibid., 51).



**Fig. 1** Basic model

This argument is the main theme of the one-third of the first half of the TMS. Knowing background information is required for completeness of sympathy. In other words, if we know the deeper background of behavior, sympathy is qualitatively deeper. It is not a unique idea that additional information changes the qualitative state of cognition. It is an important to note that an observer suspends sympathy by acquiring more information and pretends to sympathize with an objective situation. Although it is not clear what kind of influence, along with some disposition, this has on sympathy in the TMS, it maintains interest in the objective situation by evoking curiosity and encouraging deeper understanding.

The property of every passion excited by objects peculiarly related to ourselves, the pitch which the spectator can go along with, must lie, it is evident, in a certain mediocrity. If the passion is too high, or if it is too low, he cannot enter into it. Grief and resentment for private misfortunes and injuries may easily, for example, be too high, and in the greater part of mankind they are so. (ibid., 76)

This paragraph suggests that Smith did not consider the fact that an agent sympathizes equally for every person. He asserted that the degree of emotion needs to be neither too strong nor too feeble but should reach a golden mean for an individual B to sympathize efficiently. It is important to note that Smith supposed that the ability to sympathize is dependent on the state of individual B's sentiment rather than the best intensity of signals to maximize their sympathy to A. Therefore, Smith considered that individual B's acceptance of individual A's emotion is more disturbed by a sense of discomfort than the problem in this case. This process is illustrated as individual A not only sending the main signal expressing an immediate sentiment but also as a sub-signal about the way emotion is expressed. Individual B increases the rate of discount of Individual's A input.

In equal degrees of merit there is scarce any man who does not respect more the rich and the great than the poor and the humble. With most men the presumption and vanity of the former are much more admired than the real and solid merit of the latter. It is scarce agreeable to good morals, or even to good language, perhaps, to say, that mere wealth and greatness, abstracted from merit and virtue, deserve our respect. (ibid., 128)

Although it is intuitively sensible, this processing mechanism is complicated. This suggests that objective A's circumstances such as the social and economic position, cause a cognitive bias. In this case, signals about individual A's action and position

are input simultaneously and interpreted as information about individual A's state through the first transformation. We therefore assumed that the information on individual A's position influences the input or output of the second transformation. In this paragraph, although individual A's action is correctly recognized, information on their social position influences the output because "respect" is an expression of emotion dependent on individual A's position (Fig. 3). In modern cognitive science, a person takes advantage of the external order when they conceive a cognitive framework (Fiddick and Barrett 2001, Hashimoto and Egashira 2001, Lenartowicz, Weinbaum, and Braathen 2016). There are many examples in which the social order affects the activity of the mind in the TMS. Although Smith suggested that an observer's emotion is influenced by not only the behavior of an objective but also by the peripheral information, his argument does not explicitly state how information obtained is contrasted with the experience that represents an accumulation of patterns obtained in the past, and how this biases the final output. We simplified the model to treat this factor simply as a critical bias in the evocation of emotion. Although the first half of the TMS discusses several cases that influence the evocation of emotion, these are basically categorized in the cases illustrated in Figs. 2 and 3.

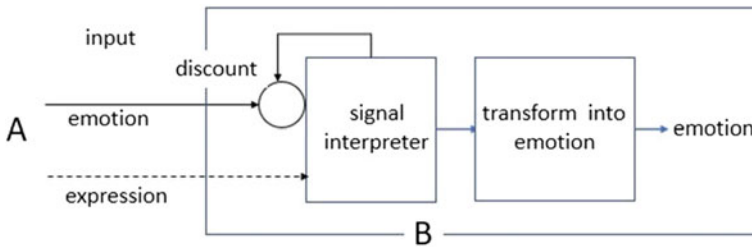


Fig. 2 Too high passion discounts an emotional input

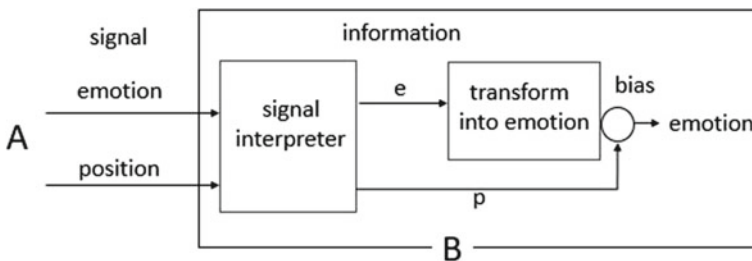


Fig. 3 Sympathy biased by social position

### 3 The Impartial Spectator

The impartial spectator, which is the most important concept in the TMS, seems to require an entirely different algorithm.

He has never dared to forget for one moment the judgment which the impartial spectator would pass upon his sentiments and conduct. He has never dared to suffer the man within the breast to be absent one moment from his attention. With the eyes of this great inmate he has always been accustomed to regard whatever relates to himself. This habit has become perfectly familiar to him: he has been in the constant practice, and, indeed, under the constant necessity, of modeling, or of endeavouring to model, not only his outward conduct and behaviour, but, as much as he can, even his inward sentiments and feelings, according to those of this awful and respectable judge. He does not merely affect the sentiments of the impartial spectator; he really adopts them. He almost identifies himself with, he almost becomes himself that impartial spectator, and scarce even feels but as that great arbiter of his conduct directs him to feel. (ibid., 206)

The concept where one imagines another person in their mind and observes the behavior of their emotions as if an outsider is complicated and ambiguous for an agent-based model that requires designing a strict procedure. Here, we considered an algorithm that could help ascertain this mechanism with complementary parts regarding areas in which Smith did not explicitly explain.

The composition of an observation from the perspective of an outsider to one who observes another person's behavior is a kind of metacognition. Recent studies elucidate the fact that metacognition plays an important role in decision-making (Ormond, Luszcz, Mann, and Beswick 1991). However, they point out that not all people necessarily have an impartial spectator as in the TMS because in modern cognitive psychology, metacognition is considered an ability that is usually found in healthy developed people.

We should therefore discuss more about the concept of the “impartial spectator.”

If he would act so as that the impartial spectator may enter into the principles of his conduct, which is what of all things he has the greatest desire to do, he must upon this, as upon all other occasions, humble the arrogance of his self-love, and bring it down to something which other men can go along with. They will indulge it so far as to allow him to be more anxious about, and to pursue with more earnest assiduity, his own happiness than of any other person. Thus far, whenever they place themselves in his situation, they will readily go along with him. (ibid., 120)

Smith asserted that the impartial spectator was not only strict but also tolerant. The degree of tolerance should be defined on a level that is acceptable to others. Smith argued on how an adequate degree of tolerance was decided based on common-sense in the following section, which explains the invasion of fair play beyond the observer's tolerance limit.

However, other people's view on common-sense is also formed in the observer's subjective view as they are not able to have another person's view. While paying attention to this point, we constructed a model of the impartial spectator (Fig. 4).

Although Smith's argument does not explicitly state whether the impartial spectator is a metacognitive or an internal process, in the process shown in Fig. 3, it is

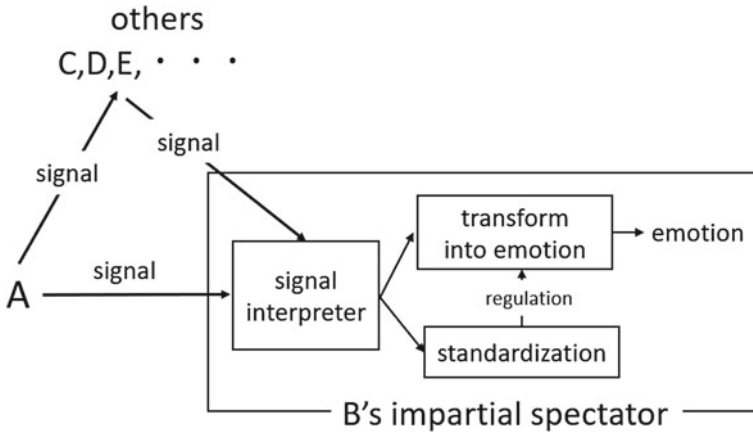


Fig. 4 The judgement process of an impartial spectator

important that the unit of the impartial spectator outputs emotion. The acceptance of the metacognitive aspect of the impartial spectator means that observer B judges A’s behavior compared with the emotion shown by both the impartial spectator and the additional process shown in Fig. 3. On the other hand, if it is not a metacognitive process, B judges according to the impartial spectator’s output. Accepting the metacognitive property of the impartial spectator raises more questions about the kind of process of comparison between the judgment of the impartial spectator and B required. In other words, this decides which judgment is finally selected, or how much weight should be attached to it. On the one hand, supposing a moral adjudicator, means that Smith’s argument assumes meta-metacognition. On the other hand, if the impartial spectator also makes this decision, it means that the impartial spectator might produce an output inconsistent with B’s judgment.

If we do not accept the metacognition account, B’s judgment completely agrees with that of the impartial spectator, which means that it is unnecessary to posit another process in B’s cognition.

This point is important in the implementation of Smith’s cognitive process as an agent-based simulation. If we do not accept meta-metacognition, the relationship between the processes of the impartial spectator and B is depicted in Fig. 5.

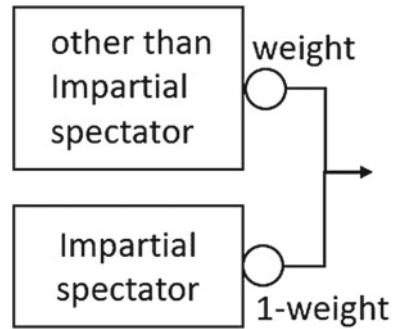
B’s final output is a composite of the emotions of the impartial spectator and B. The weights ( $0 \leq \text{weight} \leq 1$ ) of each output express which and how much the outputs are regarded as important. If the weight = 0, observer B’s conclusion completely follows the impartial spectator’s judgment and if the weight = 1, their conclusion does not follow the impartial spectator’s judgment. This logic circuit shows that it is possible that B barely follows the impartial spectator’s judgment.

On the other hand, Fig. 6 depicts the model upon accepting meta-meta-cognition.

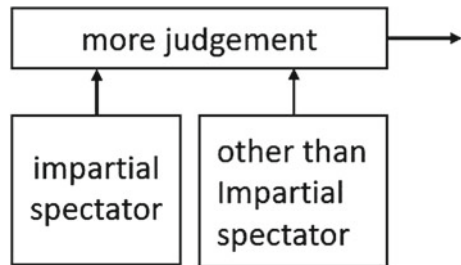
In this case, more mechanisms are required for complete judgment. The model becomes more complicated upon considering the mechanism’s formation process.



**Fig. 5** The final output process without meta-metacognition



**Fig. 6** The final output process with meta-meta-cognition



However, although Smith deemed the concept of the impartial spectator important in the basis of morality, it is difficult to consider that he supposed the meta-metacognition structure. Smith supposed that although individuals are fundamentally isolated, they possibly have a friend in their minds who does not necessarily always make healthy judgments. Moreover, Smith supposed that although the impartial spectator is a native property that every person might possess, they cannot obtain it without putting in any effort. Although Smith did not explain this, the process of developing an impartial spectator is either gradually separated from self-interest or begins to work through an accumulation of personal experiences. This argument can be applied not only to the impartial spectator but also to the general process of cognition. The accumulation of personal experience encourages one to develop a mechanism for interpreting signals and transforming them into emotions. It is not clear how Smith understood cognitive processes other than the impartial spectator. According to Smith, self-interest and selfishness are instinctual while the ability to sympathize is innate. However, adequate development of the impartial spectator requires accumulation of experience and education. Although the selfish nature of a person is enforced, the impartial spectator is separated during development and supervises its own behavior. Interactions between people who balance their self-interest with an impartial spectator are a source of morality.

### 4 An Integrated Model

Based on the preceding arguments, we designed the final cognitive model based on Smith’s theory. Smith did not demonstrate each block’s mechanisms. Here we regard them as black boxes. The internal structure of an agent was designed under this assumption as shown in Fig. 6. This model includes a part of the processing unit that is based on self-interest and one based on an impartial spectator for appreciation of a person’s behavior and emotions. The inputs and outputs of the units are adjusted by feedforward and feedback mechanisms, which are simplified in this figure, as previously explained. Therefore, Smith deemed public education and a standard of life as important to improve morality of labor class.

We did not explain the hidden layer in this model. Although this part was not described in the TMS, this system has a function that adjusts the weights of the self-interest and impartial spectator parts depending on accumulated experiences that are obtained as a result of the comparisons of the results of judgments and ex-post facts. Although we could design the model so that the difference between the final judgment and an ex-post result is provided directly to previous stages, implementing this feedback system in an agent model, should remind is that multiple loops of the same signal create the possibility of biased and unstable results. In this model, the unit of the impartial spectator is built in, and the value of the weight parameter is set to zero (Fig. 7).

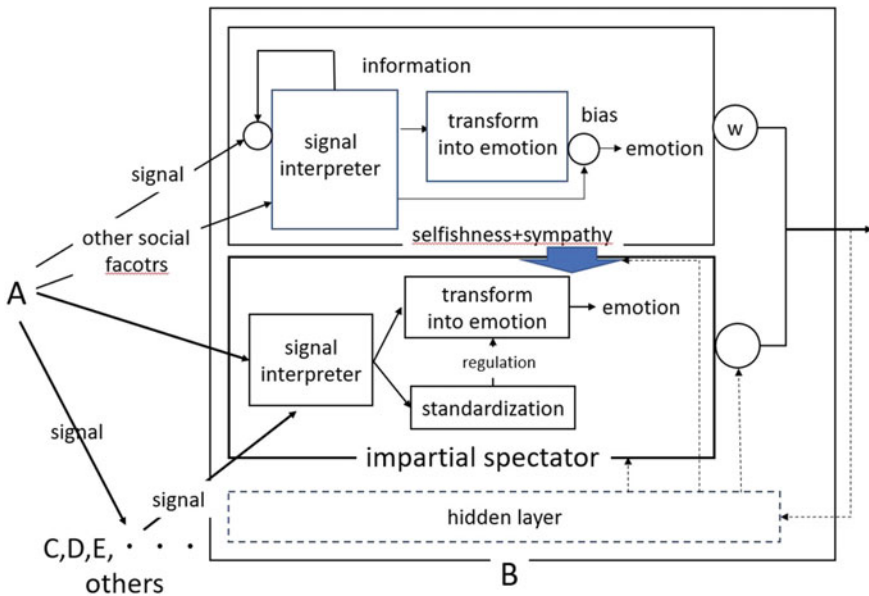


Fig. 7 A cognitive framework of Smith’s agent

Does morality emerge upon setting up many agents to form a society? When we define “morality” as the dominant norm in a society, there are two points in our model. One is an input signal to the impartial spectator that expresses the agreement of other people, and of standardization. Many preceding studies show that a social cluster is formed by referring to other people’s behaviors (Tomochi et al. 2005). If a sense of impartiality is based on the impartial spectator, it is possible for commonality to emerge from interaction of agents. On the other hand, observation by an isolated person is subjective, hence, signals are required for interpretation through a cognitive framework. Although the same interpretation requires the same cognitive frameworks, Smith did not mention any mechanism that communalizes the framework. Smith did not explicitly suggest a change in a signal interpreter or transformer or a feedback mechanism of experience because learning mechanisms or processes are not considered in the TMS. Although Smith hardly discussed this theme, he partially explained it as follows:

It is thus that the general rules of morality are formed. They are ultimately founded upon experience of what, in particular instances, our moral faculties, our natural sense of merit and propriety, approve or disapprove of. We do not originally approve or condemn particular actions because, upon examination, they appear to be agreeable or inconsistent with a certain general rule. The general rule, on the contrary, is formed by finding from experience that all actions of a certain kind, or circumstanced in a certain manner, are approved or disapproved of. (*ibid.*, 221–2).

This paragraph suggests that Smith also considered learning from experience in the process of the formation of moral sentiments. It is difficult to form a consistent set of norms for individuals without a feedback mechanism of the ex-post assessment of decisions and interaction with the external world.

Another challenge in running this agent-based system is what the agent’s incentive is. If the incentive is designed simply to pursue self-interest, the order is described as a Nash equilibrium. Alternatively, if we design a game so that agents who successfully synchronize with other agents obtain high scores, it will be possible to achieve a stable order. If we accept the hypothesis that the TMS presupposed on the arguments on self-interest in the WN, which are the foundation of the division of labor and economic growth, the mechanism of the emergence of the impartial spectator in the process of seeking self-interest and the commonality of judgment of the justification of behaviors needs to be explained.

## 5 Conclusion

The approach in this chapter is not traditional but a form of constructionism that discusses Smith’s human image as a foundation of morality. Smith did not discuss the mechanism of learning from experience, and his argument was not an ontogeny in which the spectator develops through trial and error. However, he considered that an uneducated labor class lacks a sense of morality and does not believe that a moral society only appears through self-interest. Although every person has the ability to

sympathize, it is not necessary that every person acquires the impartial spectator. Moreover, education is required for its development.

Although artificial intelligence has played an important role in our society, there are few arguments about morality in a society in which different kinds of artificial intelligence interact. In an age when free markets with virtual currency and artificial intelligence are dominant and have become a global standard, the TMS will be one of the best textbooks from which to refer morality in our society.

This chapter does not discuss Smith's argument on virtue. Morality is a general rule of behavior. However, virtue is a more favorable value. Artificial intelligence selects and improves behaviors only according to utilitarian rules. If we seek morality in our society, the concept of virtue as stated in the TMS, which is not based on utilitarianism, will suggest yet another challenge for a society with artificial intelligence.

## References

- Ashraf, N. 2005. Adam Smith, behavioral economist. *The Journal of Economic Perspective* 129 (3, Summer): 131–145.
- Fiddick, L., and H.C. Barrett. 2001. Evolution of cognition: An adaptationist perspective. In *International encyclopedia of the social & behavioral sciences*, 4996–5000.
- Harry, E., H.E. Foundalis, and M. Martínez. 2007. A generalization of Hebbian learning in perceptual and conceptual categorization. In *Proceedings of the European cognitive science conference*, Delphi, Greece, May 2007, 312–317.
- Hashimoto, T., and S. Egashira. 2001. Formation of social norms in communicating agents with cognitive frameworks. *Journal of Systems Science and Complexity* 14 (1): 54–74.
- Hayek, F.A. 1952. *The sensory order*. Illinois: Chicago University Press.
- Hebb, D.O. 1949. *The organization of behavior*. London, Routledge: Psychology Press.
- Lenartowicz, M., D. Weinbaum, and P. Braathen. 2016. The individuation of social systems: A cognitive framework. *Procedia Computer Science* 88: 15–20.
- Ormond, C., M.A. Luszcz, L. Mann, and G. Beswick. 1991. A metacognitive analysis of decision making in adolescence. *Journal of Adolescence* 14(3): 275–91.
- Smith, A. 1759 [1969]. *The theory of moral sentiments*, “New Edition” published in London in 1853. Indianapolis, Liberty Classics.
- Smith, A. 1776. *The wealth of nations*.
- Smith, A. 1983. The Glasgow edition of the works and correspondence of Adam Smith, IV. In *Lectures on Rhetoric and Belles Letters*, ed. Jo C.B. Fyfe.
- Tomochi, M., H. Murata, and M. Kono. 2005. A consumer-based model of competitive diffusion: The multiplicative effects of global and local network externalities. *Journal of Evolutionary Economics* 15 (3): 273–295. <https://doi.org/10.1007/s00191-005-0245-0>.

# The “Self” and the “Others”: From Game Theory to Behavioral and Neuroeconomics



Christian Schmidt

**Abstract** The starting point of this chapter is the evidence that the “self” is to be understood by reference to the “others”. Given this game situations offer the best framework for analyzing their dynamic relations and thus making game theory as the appropriate tool. The first part recalls the evolution of the ideas on the topic, from Leibniz studies on board games to Game theory, and from Von Neumann Morgenstern approach by means of cooperative games to Nash investigation by means of non-cooperative games. In the second part, the concept of “common knowledge” which was introduced by game theorists for linking the self to the others is analyzed from a logic view point, as well as from a mentalist approach. More precisely the assumption that rationality is common knowledge between the players is discussed and criticized. In a third part the recent contributions of neurosciences to the topic are introduced and discussed through three main dimensions: The discover of the “mirror neurons”, the contribution of the theory of mind and the neural substates of empathy. The idea of a connection between the” self” and “the others” ’by means of a kind of mental dictionary is suggested in the conclusion.

**Keywords** Common knowledge · Empathy · Game theory · Neural mirrors · Others · Self

## 1 Introduction

At first glance the understanding of the “self” is to be derived from its opposition to” the others”. Indeed, a marker of the “self” takes the form of a kind of border line between “I”(inside) and “the others” (outside). But such a prior distinction is not sufficient to capture the complete meaning of the “self”. Among the components of this outside, the “others” occupy a specific place because they are also “selves”, but from their own view point. So, if the notion of the “self” is intricately connected to its relation to “the others”, the relation between the “self” and “the others” is extremely

---

C. Schmidt (✉)

Emeritus Professor, Department of Economics, Paris Dauphine University, Paris, France  
e-mail: [christian.schmidt@dauphine.psl.eu](mailto:christian.schmidt@dauphine.psl.eu)

complex. In order to analyze and discuss the different dimensions of this complexity, we propose to use the perspective of a game situation.

The game has been chosen as an efficient support for revealing the hidden resorts of this relation. Indeed, a game refers to situations where the “self” of each player is directly related with the “self” of the other players. Even in the case of games of pure chance, the players are supposed to play with some “other”, outside of themselves, which takes the form of hazard (or chance). Starting from the study of board games, Leibniz had already formulated the same idea with his own words in the seventeenth century (Leibniz, 1676). Therefore, Leibniz developments on games could retrospectively be understood as an early step in the direction of Game Theory by means of a Theory of games (1).

The nature of the relations between “myself” and the other “selves” is a key question in microeconomics, and more generally, in social sciences. Game theory has provided an analytical format for modeling the interactive process of decision-making between individual agents in almost all kinds of social interactions. It will therefore be used as a starting point for our quest. Let us remember however that Game theory has been preceded by a tentative mathematical theory of board games developed during the first part of the twentieth century mainly by several French mathematicians, as Borel and Ville (1938) and de Possel (1936). For them, board games are characterized by three main components: Self (the player), the other(s) player(s), and the game which refers to a set of explicit rules known by all the players. If the self is the referential point for each player when he chooses his move, the chosen move is also the result of a self-reaction to the last moves chosen by the other player (s) as well as the expectation of his next moves (see: Bridge, chess, Go...). As for the rules of the game, they have a specific role in the process of playing. They work as a kind of system which organize the meeting between the “self” and the “other selves” during the play, which is an additional argument justifying Leibniz interest for boardgames. Let us note that a revival interest in the board games studies and researches has recently taken place, thanks in particular to the new analytical tools developed by the Neurosciences (Gobet et al. 2004).

We propose to enlarge and deepen the analysis of this connection between the “self” and the “others” in the game theoretical format. Of course, von Neumann, and Morgenstern, and after them Nash, have proposed mathematical solutions to board games according to their specific rules (Poker, chess, ..., and even Hex for Nash) (2). But Game Theory has a larger and more ambitious topic: To propose a general theory to be applied to all kinds of interactive situations whatever their specificities, and so, whatever their rules. De facto the implicit rules of the games correspond to the properties of the agreements between the players during the play: Either the agreements are binding, the game is call cooperative and the players can cooperate by means of alliances; either the agreements are not binding, the game is call non-cooperative because the players cannot build real alliances. The categories of cooperative and non-cooperative games correspond to two types of interactive social situations. Their meaning does not necessarily refer to realistic situations. They must be understood as hypothetical assumptions relevant for game theoretical modeling.

This means that the reference to precise rules of the games which are determinant in almost board games is not so significant in Game Theory.

The interaction between the players is analyzed in game theory as a kind of mechanism relating directly the decision-making of each player to the other players by means of the assumption of rationality in its economic definition. So, the decision of each player is driven by his “self” (self-interest maximization) and each player is supposed to know that it is the same for the other players. Therefore, rationality has been assumed to be “common knowledge” between the players of the game. Thanks to this assumption, each player is supposed to be able to put himself in the other’s shoes, and then to predict the decision of the other players of a social game by means of the hypothesis that other players are “other-selves” as themselves.

The first part of this chapter is devoted to a critical re-examination of the explicit, and sometime implicit, assumptions used by the classical game theory for modeling the complex relation between the self and the others during all kinds of interactive processes. But the game theory approach to such interactions of the question raises more fundamental questions. If the players’ decisions are taken from the viewpoint of each player “self”, then, what a player really knows about the self of another player when common knowledge rationality is questioned? The question is to be extended to the game theorists themselves, who being not in the same position as the players vis a vis the game, do not necessarily apprehend their interaction in the same way that the players do. In a stimulated paper Aumann and Brandenburger wrote in a provocative fashion: “belief systems are primarily a convenient framework to enable us, -the analysts- to discuss the things we want to discuss: actions, payoffs, beliefs, rationality, equilibrium, and so on. As for the players themselves, it’s not clear that they need concern themselves with the structure of the model...” (Aumann, Brandenburger 1995).

In order to take this difficulty into account, some economists and philosophers have proposed to build appropriate models derived from epistemic logic (Bacharach, Gerard-Varet, Mongin and Shin 1997) (3). So, they explicitly introduce the logical supports of the notions of knowledge and belief for taking into account, (1) the game theoreticians knowledge and the beliefs of each players on the others players’ knowledge and belief during the interactive decision process of the game, (2) the knowledge of the game theoreticians on the knowledge and the beliefs of the players of the game. Such an approach opens the way to a multilevel of beliefs for understanding what the other players’ decisions can really mean from the “self-viewpoint” of each player (Aumann 1999; Fagin et al. 1999). Indeed, others ‘selves do not belong to a knowledge commonly shared by all the players (and the game theoreticians) but to their “own world” that the other players can only apprehend by means of a kind of code to be translated into their personal world.

The second part of the chapter analyses the contributions, but also the limits of such kinds of epistemic logics for improving the understanding of the relation between the “self” and the “others” during the interactive processes. Rebuilding a relevant logic system is a necessary but not a sufficient step for deepening the investigation of this relation. It mainly concerns the analysts from the outside, but does not really

provide information about the mental representations of the “self” and the “others” by the subjects themselves.

Fortunately, in mixing the data provided by experimental economics with neural information derived from neuroscientific researches, new elements have emerged which are introduced and discussed in a third part. First of all, many well-known experiments reveal that the rational choices of individual subjects do not necessarily coincide with a maximization of their self-interest (Camerer 2003). In addition, two different notions have emerged from recent neuroscientific works: On one hand the empathic relation linking the “own self” to the “other selves”, on the other hand the possibility to attribute mental states to others according to the “theory of mind (TOM). They take the form of interpersonal communications system between the “self” and “the other”. Their neural origin has been identified and their overlapping start to be recognized. Indeed, the empathic system primarily activates the emotive functions of the brain, but it can also concern some related cognitive functions, such as decision in interactive contexts and the sense of social morality in case of social choices (Decety 2011, 2015). Several questions remain open concerning the nature of their connections during the process of mentalization

The conclusion re-evaluates the concept of “self” in its relation to “the others”, in the light of the various contributions which are analyzed. It proposes a new acceptance of what is call “common” to the self and the others during their interaction.

## **2 Self-interest Players’ Interactions in Game Theoretical Situations from Von Neumann Morgenstern to Nash Perspective**

The relations between the “self” and the “others” in game theory takes the following form: How to be computed the result(s) of the interactions (the game) between self-interest decisions taken by individual agents (the players)? If the assumption of players’ self-interest is decisive (the definition of self-interest being understood in its economic meaning of utility maximization). Let us recall that von Neuman introduced one of the first formal definition of the utility maximization as a criterion of rationality (Von Neumann 1928). So, each player of the game is supposed to choose her strategy by maximizing her own expected utility. But in spite of a common reference to individual rationality, the supposed relation between the self and the others is not understood in the same way in cooperative games studied by von Neumann, Morgenstern and their successors, and in non-cooperative games promoted by Nash.



## ***2.1 The Social Reference to a Common “Standard of Behavior” in the Initial Cooperative Game Format***

In the first edition of “Game Theory and Economic Behavior” the two authors insisted to precise that they consider the problem of rational behavior from an altogether different angle than the economists of their time, due to their perspective of a “game theory”, where the players are the participants of “social games of strategy” (Von Neumann and Morgenstern 1944). Then, they explained that the starting point is not to be found in a pure individualistic approach of rationality corresponding to a Robinson Crusoe type, but rather into a set of rules for each player to be derived from the definition of a game solution. Therefore, the game in its mathematical treatment is considered as the key concept. More precisely, the solution of a game situation in its mathematical formulation can be understood as providing “A set of rules for each participant which tell him how to behave in every situation which may conceivable arise” (Von Neumann and Morgenstern 1944, p. 31). For von Neumann and Morgenstern such an approach to the rational behavior is closely related to the specificity of game situations directly derived from board games examples where the rules of the games are supposed to be known by all the participants (Chess, Poker...) and can be extended to economic relations and social organizations. Therefore, the rationality of the game players is not defined in terms of individual preferences, but rather in terms of collective preferences by reference to a social order.

However, such a definition of rationality raises questions when it is applied to individual players’ decision. Indeed, the rules of rational behavior so defined for each individual player must take into account all kinds of behaviors of the other players, including irrational behaviors. But how a rational player according to the rules previously quoted could be capable of expecting the irrational behaviors of the other players? For von Neumann and Morgenstern however such a question is not relevant and its answer is not to be found at the level of each individual player of the game, but rather at the collective level of the game solution itself, where individual players are taken into accounts as simple components of the game. Then, the rules of rationality deduced from the solution of the game is to be understood in terms of what they call the “standards of behavior” for the players of the game. More precisely, the solution of the game can be described as a system of imputations which refers to an “established order of society” from a socioeconomic viewpoint corresponding to “accepted standards of behavior” for the players. In the end, according to the authors, the foundations of such systems are based on a mathematical formulation of the social stability (Von Neumann and Morgenstern 1944, pp. 40–45).

Such an interpretation of players’ individual behaviors in interactions remains questionable for different reasons. First, in many cases several solution concepts have been later proposed by different game theorists (The Core, the Stable Set(s), the Shapley value...), which can be associated to the same game situation. It means the possibility of alternative systems of imputations corresponding to different standards of behavior. Secondly, and still more disputable, according to this view the real subjects of the imputations are no longer the individual players themselves, but the

alliances (or coalitions) that the players can build. Consequently, there is no room for the “self” of the players in this first version of a general theory of games. The domain of the Game Theory initially proposed by the Von Neuman and Morgenstern founding book is thus de facto limited to what was call later the “cooperative games”, as soon as we leave the very specific case of the two persons zero-sum games.

## ***2.2 The Emergence of the “Self” and “the Other” in the Non-cooperative Formulation of a Game Situation***

A quite different approach to games has been introduced by Nash from his PhD dissertation, under the simple title of “Non-cooperative game” (Nash 1950). As specified by Nash, the distinction between Cooperative and non-cooperative games is not the result of a mathematical treatments of game situations, but rather the consequence of their initial assumptions. Whilst the coalitions formed by the players is the main topic of the von Neumann and Morgenstern’s theory of games, Nash starts from an opposite perspective where there is no coalitions and where “each participant acts independently, without collaboration or communication with any of the others” (Nash 1951). Such a different perspective for studying interactive game situations re-introduce the individual subjects in the center of the analysis.

One first consequence of the new paradigm of non-cooperative game proposed by Nash concerns the understanding of players’ rational behavior. Without reference to coalitions, there is no longer possibility to infer the rationality from the social imputations associated to a solution concept. Therefore, Nash comes back to individual rationality defined in terms of economic utility maximization, first introduced by von Neumann and Morgenstern and then developed and specified by Savage (1953). However, due to the interactive specific characteristics of games situations, the rational behavior of each player is now directly dependent of his expectations about the rational behaviors of the other players. For Nash, a mathematical solution can be found by means of his concept of equilibrium, by assuming that all the players are supposed to perfectly know what he calls “the full structure of the game”. We can recognize the well-known idea that a game can be understood as a kind of machine for computing the best solution of interactive systems. So, all the players of board games, as chess and poker, are supposed to perfectly know their rules and the full structure of the game (5). But the full structure of a game situation is much more complex and cannot be reduced to its formal rules as in boardgames. Nash himself admits that his mathematical model is to be considered much more as a useful heuristic rather than a realistic solution. (Nash appendix 1950a).

What does really mean here “a perfect knowledge of the full structure of the game” for the players? In the game theory approach of rational choice, the logical framework of a game looks like the support used by Savage for modeling the rational choice of individual decision-making. For Savage, the “world” of the rational choice is what he calls a “small world”, from which the sets of the world are to be chosen by

the individual decision-maker (Savage 1954). However, one can legitimately wonder if a game situation can really be assimilated to a “small world” à la Savage? Indeed, the matrix of a  $2 \times 2$  non-cooperative game takes the same form as a table of the expected utility theory for an individual. But in the “Savage small world”, the states of this world are supposed to be out of the control of the decision maker and outcomes are supposed to be the direct consequences of the chosen actions on each state. In the world of a game, the states are for each individual player directly dependent of the decision taken by the other players of the game, including himself. Therefore, the solution proposed by Nash for solving non-cooperative games, the well-known Nash equilibrium, cannot be reduced to a simple extension of the expected utility theory to two and more individuals. It questions the supposed “world of the game”, and its knowledge (and even its representation) by each of the players. This is a much more difficult question which cannot be reduced to a simple distance often observed between a theoretical solution and its empirical results, as Nash has quoted in the end of his paper.

The interactive process in a non-cooperative game situation transforms players’ in two ways:

- (1) The expected outcomes cannot be just deduced by each individual player from the consequences of her chosen action, because the values of the outcomes also depend to the actions chosen by the other players. As a result, each player has to take into account a second level of expectation on the expectations of the other players. But the problem is still more complicated, because their expectations are also dependent of her own expectation and so on..., and those levels of expectations are not independent but closely interlinked. In brief, players’ expectations of a non-cooperative game take the form of an intricate multilevel and non-closed system.
- (2) The reference to individual rationality à la Savage is not sufficient to support the players’ expectations so pictured. Indeed, players’ rationality refers to their own preferences and each player does not know a priori the preferences of the other players. In conclusion, players expectations which lead their rational decision-making in a non-cooperative game situation must take into account their self-interest, but also the supposed self-interest of the other players as well as the multilevel structure of their inter-dependence.

### **3 The Rationality Introduced as a Common Knowledge Between the “Self” and “the Others” in Game Situations**

In order to solve those difficulties, game theorists have introduced (Aumann 1976) and developed during the years 1980s and 1990s, the additional assumption that not only each player of a non-cooperative game is supposed to be rational (in the sense of a self-maximiser), but, in addition, that rationality must be a common knowledge between the players for reaching if the solution concept introduced by Nah for non-cooperative games is to be reached.

The idea of a “common Knowledge” was initially be borrowed from Lewis who used it to solve the coordination problem of mutual expectations raised by human conventions (Lewis 1969). A highway code chosen by Lewis provides a good illustration of its meaning to be applied to mutual expectations. Indeed, each driver is supposed to know the highway code, that the others also know it, and that they know that she knows it, and so one...Thanks to the common reference to the highway code, the belief of each driver about the knowledge of the highway code knowledge of the other drivers is transformed in a knowledge which becomes a common knowledge as the different levels of beliefs tend to infinity (4).

Out of its logical epistemic dimensions, the common knowledge, which allows in this example to coordinate the drivers’ mutual expectations, is the knowledge of an objective and well formulated unique code of behavior for the drivers, namely the highway code. The economists’ purpose was to extend the social property of a common knowledge to the rationality which is assumed to be shared by all the players, and then by the self and the other selves in the game situation. But such a transposition of the abstract notion of rationality into a reference code of behavior for all players in a game situation -which, could allow each individual player to forecast the strategies to be chosen by all the others- is not self- evident and raises a lot of difficult questions that need to be analyzed now. In particular, it questions the actual meaning of rationality assumed to be a common knowledge for the players of a game

### ***3.1 Can Rationality Really Be a Common Knowledge for the Players of a Non-cooperative Game?***

We have already shown that players’ rationality is classically understood in game theory as the maximization of their expected utility in line with the economic approach of individual decision-making. So, it is, at least implicitly, assumed that the players of a game act rationally according to this definition. Can we however infer from this assumption that rationality is to be understood as a set of rules of behavior known and accepted by all the players, and so must be considered as common knowledge in the meaning previously introduced by Lewis?

Indeed, it looks obvious to everybody that a positive monetary outcome corresponding to gains is rationally preferred to a negative monetary outcome corresponding to losses, and that larger gains are preferred to smaller, and the reverse for the losses. But, as soon as what is expected by each player is a subjective expected reward, not necessarily measured in monetary outcomes, its subjective valuation by each individual player could require different processes from a player to another according to his preferences, all of them being rational from their own subjective view point. Therefore, to induce the rational behavior of another player from his own approach of rational choice in the same situation looks, at first glance, to be oversimplified and even perhaps wrong.

We must recall that the maximization of the expected utility is not the only criterion of rational choice to be used by supposed rational players. Indeed, the well-known theorem demonstrated by Von Neuman to provide a game solution were closely link to the maximization of expected utility (see Maximin = Minimax in two persons zero sum games). But other criteria of rationality can be used by rational players according to their own approach of the different game situations, as for example, the minimax risk in case of risk aversion, or the minimax regret in case of self-anticipated regret, etc.... Various results of recent experimental games provide a lot of examples illustrating such diversity of rational behaviors. Incidentally, the assumption of rationality common knowledge for the players has been extended to other criterium of rationality than the maximization of expected utility, as, for example, the Minimax-regret (Hyafil and Boutilier 2004) and the Maximin (Aghassi and Bertsimas 2006), at the obvious condition however that all the players refer to the same criterion.

Out of the rational criteria, alternative data and their used by the players to compute their valuation can be different. For example, expected utility is traditionally computed on the basis of the expected outcomes of the players, but it can also be computed on other bases, such as, for example, the difference between the expected outcome and a referential outcome chosen by the players (for instance, but non exclusively, their present outcome) as it is in the prospect theory (Khaneman and Tversky 1979; Tversky and Khaneman 1992). From a positive viewpoint, a large spectrum of personal accounting methods has been more recently identified and grouped under the general term of “mental accounting” introduced by Thaler (1999, 2008).

Therefore, rationality cannot be understood as a common knowledge for the players of a game in the same way that it is the highway code in Lewis example. We argue that an abstract definition of rationality whatever cannot be transformed into a common knowledge for the players of a game for the reasons above elaborated. So, even if the assumption of rationality as a common knowledge helps us to get around the epistemic difficulty of the multilevel of knowledge which characterizes the decision process in a game situation, it does not solve its treatment for the players of the game themselves.

### ***3.2 Beyond Rationality the Contribution of a Common Belief for the Understanding of the Self and the Others in Game Situations***

In spite of its limitations, already noted by Aumann, the introduction of the logical concept of common knowledge in game theory remains decisive for the theory of non-cooperative games because it provides a logical foundation for the existence of a Nash equilibrium, as well as all its refinements (Halpern and Moses 2017). Indeed, if rationality in terms of maximization of the expected outcomes by the players is common knowledge, it follows logically that there exists at least one Nash equilibrium in all the correspondent game situations. We have seen that such a result

has been extended to other solution concepts derived from Nash equilibrium and based on alternative definitions of rationality. So, according to Aumann position, the assumption of rationality common knowledge as a theoretical hypothesis more concerns the theoreticians of game than the players of the game themselves, who are free to do what they will. More recently, Aumann and his colleagues have demonstrated that even rationality common knowledge was not necessary for an epistemic foundation of a Nash equilibrium, which just requires a common belief, and even a mutual belief in case of two players, because the multilevel of belief and knowledge are only required for more than two players (Aumann and Brandenburger 2016).

Indeed, this epistemic approach to games as a belief system opens the way to new theoretical perspectives, but it also concerns the players for other reasons. The key problem for the players is no more the rationality but the belief of the others' decision, which can be summarized as follows: "I believe that the other will..., as well as the other player believes that I shall..." "If this belief is certainty, then the belief becomes a knowledge. Common rationality could obviously be a support of such a common knowledge, but other mind faculties of human individuals also contribute to rely "myself" and "the others" in a common belief. Recent experimental results and the investigation of their corresponding neural activities now helps us to better understand their role in the functioning of the belief system behind the interactive situations studies by game theory. They will be surveyed now.

#### **4 The Others" as a Necessary Mental Complement of the "Self" for the Players**

Several well-known experimental behavioral games have shown that thoughtful players do not choose the strategy corresponding to the maximization of their self-interest. This has been particularly observed in the dictator, the ultimatum and the trust games, but also in many other experimental games (Camerer 2003). In the ultimatum game, which is probably the more popular, the first player, who proposes the sharing rule for a given amount of money to the other player, does not optimize his own interest and both players agree on a non-Nash equilibrium. A first economic interpretation would be to impute this result to irrationality of the players. But we have shown the limitations of the reference to rationality in the interactive situations. A more sociological explanation might refer to a social code followed by the two players. But social codes are different from a culture to another, and however those players' behaviors have been observed almost everywhere in the world and in spite of significant cultural differences (Heinrich 2000; Heinrich et al. 2001) (5). Therefore, we propose another explanation of those experimental results in line with our analysis of the common belief in interactive situations, but taking into account what we start to know about players' brain functioning.

#### ***4.1 Three Possible Approaches of a Common Belief Linking Each “Self” to “Other Selves” Derived from Neurosciences***

A first element can be found in the discovery of a neural system specifically devoted to directly translating the observation of moves realized by others into the activations of the same part of the observer’s body. The corresponding neurons that were discovered are called the “mirror neurons”. They were initially pointed out in some species of monkeys, and were rapidly extended to humans (Rizzolati et al. 1996; Rizzolati and Croighero 2004). Its domain has been enlarged to the intentions (intentional stances), revealing the existence of a direct connection between the “server” and the “receiver” working beyond their cognitive understanding (Jacoboni et al. 2005). This neural connection can be viewed as a common support of communication related to the notion of mutual belief previously introduced. Nevertheless, in spite of its importance, the mirror neurons domain remains limited. It suggests however that the mental relation between the “self” and “the others” starts at the level of a kind of “intercorporeality” (Fuchs 2017).

Another property of the human brain has been also introduced, firstly at an experimental level, which is now considered as a main component of the Theory of Minds, “The mind reading.” According to it, the human brain of individuals is able to assess the states of the mind of others, just by looking at their faces (Gallagher and Frith 2003). Such brain capacities are relevant for our topic. Applications to games have shown the mental differences for a player to play against other players and to play against a computer (Gallagher et al. 2002). In addition, further functional magnetic studies (fMRI) studies applied to board games, and especially to chess game, have revealed that chess players mobilize almost the same neural networks than those which have been identified for discriminate different levels of knowledge in the theory of mind (Powell et al. 2017). Therefore, it opens the way to a better understanding of the neural mechanisms behind the notions of “mutual” and “common belief” previously introduced in Game Theory. Different levels of knowledge or beliefs are indeed to be distinguished in the brain cognitive activity. In a first step, the states of mind are modelled by the brain, and then are attributed to the self, and (or) to the others in a second step. Until now, the fMRI investigations have just allowed us to identify two different neural networks thanks to the attribution of the states of mind to the “self” and to “the others”. But in part of their clear distinction, a closed connection has also been detected in their functioning (Voegeley et al. 2001). The dynamics of their complementarity seems to be a key for the understanding of all those forms of intersubjectivity (Voegeley 2017).

Both fields of research mentioned primarily concern the cognitive dimension of the intersubjectivity, but there is another dimension, perhaps still more obvious, that also concerns the intersubjectivity, the emotional dimension. The term empathy, which must not be confused with sympathy and which is not quite antagonistic to antipathy for the same reason, captures its meaning. Strictly speaking, empathy designates the faculty of sharing an emotion felt by another person. But such a faculty also requires



the representation of the other feeling, which implies a cognitive perspective. As for the other components of the intersubjectivity previously introduced, the neural reference of empathy starts to be known, thanks in particular to the fMRI contributions. But the most interesting is its possible relation with social justice, which becomes to be studied by some researchers as Decety and Cowell (2014), Cowell and Decety (2015). This connection could explain, for instance, the results repeatedly observed in the ultimatum game. The decision of the first player seems to be inspired by a mixed feeling of empathy and morality. In addition, he supposes that this feeling is also shared by the other player, who really accepts the rule proposed by the first player. Empathy so understood could be a foundation for a kind of mutual belief, which leads to the strategies chosen by the both players in the game situations.

#### ***4.2 Toward the Understanding of an Interactive Dynamics Linking the “Self” to the Other “Selves”***

The neural components of the different dimensions relating the self to the other selves previously identified above are not sufficient for explaining how they are combined in interactive situations. Thanks to the development of the Theory of Mind, we know now the two different components corresponding to the cognitive operations, firstly to identify a state of mind, and then to attribute this state to another, which follow different cognitive circuits in the brain. But such information does not say anything about the relations between those operations and others mental operations pushed by the empathic dispositions of the brain.

First of all, by using a novel method, a recent experiment has revealed that, in spite of the closed and embedded relations between the cognitive properties of the Theory of Mind (TOM) and the emotional dimension of the Empathic system, each of them activates, different kinds of neural networks (Kanske et al. 2015). In order to understand the functioning of their relation it is necessary to capture the modalities of their interactions during the dynamic of their interactive meeting. Therefore, additional studies are necessary for identifying the interactive mental processes between the self and the others all along the different steps of a game situation. Here also, the reference to games is illustrative.

Two types of social interactions have been distinguished according to their different purposes, namely cooperative and competitive. A first study succeeded to identify their neural correlates by means of an fMRI investigation where players are supposed to play in computer games where the rules are cooperative in the first case and competitive in the second case. Let us note the direct relation with the cooperative and noncooperative game in game theory and the introduction of a common knowledge by means of a social code (the rules of the game) proposed by Lewis. This study reveals some common neural activations, cognitive as well as emotional in players' brains, and some different activations corresponding to the different rules of each type of game (Decety et al. 2004)). As such a general framework of brain



working has been observed with small differences in all the participants, we can suppose that the same activated brain areas tend to support a kind of common belief derived from the rules of the game in both cases.

In this experiment however, there is no direct personal interactions because each player does not play with another player, but with computer games. In addition, such an experimental design does not take into account the whole dimensions of a real game situation, which involves expectations and memory. Fortunately, a very recent experiment using a new technic of “hyperscanning” has succeeded to capture the brain interactions of the players during the course of this two types of game (Spilakova et al. 2019). This new technic of brain investigations opens interesting perspectives which concern how really works the dynamic relations between the emotional empathic dimension and the cognitive reasoning dimension of the relations between the self and the others. Once again, such a line of research meets several neural investigations concerning games, and now board games, where similar kinds of neural activated combinations have been observed, contrasting for example, defensive an offensive strategy chosen by the players during the development of different games, as Shogi, a Japanese version of Chess game (Wan et al. 2015; Kolling and Hunt 2015). Further analyses are still necessary to confirm and enlarged those interesting first results.

## 5 Conclusion

The recent progress of our knowledge concerning how the brain works in interactive situations has begun to improve the understanding of the complex relations between the self and the others. Several neural systems have been discovered which directly connect our mind to the mind of the others, through different identified channels and by means of distinct processes, as, for instance, the neural mirrors, the Theory of Mind and the empathic networks. But the revelation of such a human mental equipment is not sufficient for the complete understanding of the topic.

The real challenge concerns the dynamics of the interplay between different persons by means to those mental capacities which characterize the interdependence of game situations. The contributions of the neurosciences which have been used conduct to re-examine the idea of common knowledge as it was introduced by game theorists. First of all, common Knowledge cannot be considered as a data shared by the players and given once forever. Secondly, as it has been recognized by Aumann and other game theorists, that rationality whatever its meaning cannot be considered as common knowledge between the players. Nevertheless, something common, or at least mutual, relays the brain abilities of each player to the other players. This property named “common” (or “mutual “in case of two persons) is to be understood more as a kind of mental dictionary which allows to translate the signs transmitted by one player to another one. But the connection between the “self” and “the others” so realized is rather dynamic than static, and therefore enriched thanks to new information brought the duration of the game.

Let us note the key role of game situations for revealing the reality of the complex relations between the self and the others in all types of social life which implies a type of social referential code based on brain faculties even if this reference does not eliminate uncertainty (Schmidt and Livet 2014). Such a brain property could operate as a mental dictionary opening the way to a kind of translation from one self to another. Therefore, closed relations between game theorists (including those who work on boardgames) and social neuroscientists (including those who works on economic questions) seem quite necessary for improving the knowledge of the important topic reviewed in this chapter.

### Notes

- (1) Leibniz during his stay in Paris from 1672 to 1676 took the opportunity to observe the players of several boardgames of that time, as the Quinquenove, the Basette, ... According to twenty one manuscripts written during this period, the analysis of such games was the starting point of his contributions to the birth of probabilities calculation and to a real mathematical theory of games.
- (2) The game of Hex is a  $n \times n$  board game invented (or restated) by Nash in 1948, when he was a student in the mathematics department of Princeton university.
- (3) The authors of the collective book propose to revise the logical foundations of game theory in the light of the epistemic logic developed by Hintikka and Kripkte.
- (4) In the Lewis example, the common knowledge of the drivers does not concern rationality but the highway code.
- (5) With the singular exception of the Machinguenga tribe in Peruvian Amazon, whose members play as Nashian selfish interested players (Henrich, 2.000).

### References

- Aghassi, M., and D. Bertsimas. 2006. Robust game. *Mathematical programming*.
- Aumann, R.J. 1976 Agreeing to disagree. *Annals of Statistics*.
- Aumann, R.J. 1999. Interactive epistemology: Knowledge. *International Journal of Game Theory*.
- Aumann, R.J., and A. Brandenberger. (2016). Epistemic solution for Nash equilibrium. *Reading in Formal Epistemology*.
- Bacharach, M., L.A. Gerard-Varet, P. Mongin, and H. Shin. (eds). 1997. Epistemic logic and the theory of games and decision. *Kluwer*.
- Borel, E., Ville, J. 1938. *Traité des probabilités: applications aux jeux de hasard*. T.4, Gauthier-Villars.
- Camerer, C.F. 2003. *Behavioral game theory*. Princeton University Press.
- Cowell, J.M., and J. Decety. 2015. *Empathy, justice and moral behavior*. Taylor and Francis.
- Decety, J. 2011. Dissecting a neural mechanism mediating empathy. *Emotion Review*.
- Decety, J., and J.M. Cowell. 2014. The complex relation between morality and empathy. *Trends in Cognitive Sciences*.
- Fagin, Geanakoplos, and Yardi Halpern. 1999. A hierarchical approach to modeling knowledge and common knowledge. *International Journal of Game Theory*.
- Fuchs, T. 2017. *Collective body memories*. MIT Press.

- Gallagher, H.L., and D.C. Fritz. 2003. Functional imaging of “Theory of mind”. *Trends in Cognitive Sciences*.
- Gallagher, H.L., A.L., Jack, H. Roepstorff, and C.D. Fritz. 2001. Imaging the intentional stance in a competitive game. *Neuroimaging*.
- Gobet, F., J. Retschinski, and A. de Voogt. 2004. *Moves in mind: The psychology of boardgames*. Psychology Press.
- Halpern, J.Y., and Y. Moses. 2017. Characterizing solution concepts in terms of common knowledge. *International Game theory*.
- Henrich, J. 2000. Does culture matter in economic behavior? Ultimatum game bargaining among Ghemachinguenga of the Peruvian Amazon. *American Economic Review*.
- Henrich, J., R., Boyd, S., Bowles, C., Camerer, E., Fehr, H. Gintis, and R. Mac Elreath. 2001. In search of homo economicus, behavioral experiments in 15 small scale societies. *American Economic Review*.
- Hyafil, N., and C. Boutilier. 2004. Regret minimizing equilibria and mechanism for games and strict type uncertainty. In *20<sup>th</sup> conference on uncertainty*.
- Jacoboni, M., I. Molnar- Szakacs, V., Gallese, G., Baccino, J.-C. Mazziotta, and G. Rizzolatti. 2005. Grasping the intentions of others ones’ own mirror neuron system. *Plos*.
- Kanske, P., A., Böckler, F.M. Trautwein, and T. Singer. 2015. Dissecting the social brain: Introducing EmpToM to reveal distinct neural networks and brain-behavior relations for empathy and theory of mind. *Neuroimage*.
- Khaneman, D., and A. Tverski. 1979. Prospect theory, an analysis of decision under risk. *Journal of Risk and Uncertainty*.
- Kolling, N., and L.T. Hunt. 2015. Divide and conquers: Strategic decision areas. *Nature*.
- Leibniz, G.W. 1676. De numero jactum in tesseris , Sur le calcul des parties. In *L’Estime des apparences*, ed. M. Parmentier. *Vrin, Paris*.
- Lewis, D. 1969. *Convention*. Harvard University Press.
- Nash. 1950. *Non-cooperative games*. Ph.D. Princeton University.
- Nash, J. 1951. Non -cooperative games. *Annals of Mathematics*.
- de Possel, R. 1936. Sur la théorie mathématique des jeux de hasard et de réflexion. *Entretiens de Monaco*.
- Powell, J.L., D., Grossi, R. Corcona, and F. Gobet. 2017. The neural correlates of theory of mind and their role during empathy in the game of chess: A functional magnetic resonance study. *Neurosciences*.
- Rizzolatti, G., Fadiga, L., V. Gallese, and L. Fogassi. 1996. Action recognition in premotor cortex. *Brain*.
- Rizzolatti, G., and F. Graighero. 2004. The mirror-neuron system. *Annals of Neurosciences*.
- Savage, L.J. 1954. *The foundations of statistics*. Dover Press.
- Schmidt, C., and P. Livet. 2014. Comprendre nos interactions sociales, une perspective neuroéconomique. Odile Jacob.
- Spilakova, B., D.J., Shaw, K. Czekoova, and M. Brazdil. 2019. Dissection social interaction: Dual-fMRI reveals patterns of interpersonal brain- behavior relationships that dissociate among dimensions of social exchanges. *Social Cognitive and Affective Neurosciences*.
- Thaler, R.H. 1999. Mental accounting matters. *Journal of Behavioral Decision-Making*.
- Thaler, R.H. 2008. Mental accounting and consumer choice. *Marketing Science*.
- Tverski, A., and D. Khaneman. 1992. Advance in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*.
- Vogeley, K. 2017. Neural mechanisms of intersubjectivity. *Transactions of Royal Society*. Vogeley, K., Bussfeld, P., A. Newen, and S. Hermann. 2001. Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage*.
- Von Neumann, J. 1928. Die Zerlegung eines intervalles in abzählbar viele kongruente tellmengen. *Fundamenta Mathematicae*.

- Von Neumann, J., and O. Morgenstern. 1944. *Theory of game and economic behavior*. Princeton University Press.
- Wan, X., K. Cheng, and K. Tanaka. 2015. Natural encoding of opposing strategy values in anterior and posterior cortex. *Nature Neurosciences*.

# Why Is Behavioral Game Theory a Game for Economists? The Concept of Beliefs in Equilibrium



Michiru Nagatsu and Chiara Lisciandra

**Abstract** The interdisciplinary exchange between economists and psychologists has so far been more active and fruitful in the modifications of Expected Utility Theory than in those of Game Theory. We argue that this asymmetry may be explained by economists' specific way of doing equilibrium analysis of aggregate-level outcomes in their practice, and by psychologists' reluctance to fully engage with such practice. We focus on the notion of belief that is embedded in economists' practice of equilibrium analysis, more specifically Nash equilibrium, and argue that its difference from the psychological counterpart is one of the factors that makes interdisciplinary exchange in behavioral game theory more difficult.

## 1 Introduction

One of the most influential texts published in the behavioral and social sciences in the first half of the twentieth century was, according to many, Von Neumann and Morgenstern's (vNM) *Theory of Games and Economic Behavior* (TGEB hereafter) (1944). Not only did the book lay the foundation of game theory, which has become the essential research tool in contemporary economics, it also influenced several other disciplines beyond economics, from political science to linguistics and biology.

Two of the most important contributions of TGEB to economics are the axiomatic derivation of expected utility (in the second edition of 1947) and the minimax solution to zero-sum games. John Nash generalized vNM's existence proof of equilibrium in non-zero-sum games in 1951, thereby providing game theory with a foundational solution concept. Leonard Savage completed the axiomatization of expected utility

---

We thank Wade Hands, one of the editors, for valuable comments.  
Both authors equally contributed to research and writing of the article.

---

M. Nagatsu (✉)  
University of Helsinki, Helsinki, Finland  
e-mail: [michiru.nagatsu@helsinki.fi](mailto:michiru.nagatsu@helsinki.fi)

C. Lisciandra  
University of Groningen, Groningen, The Netherlands

theory with subjective probabilities in 1954, thereby founding the standard formal approach to decision-making under risk.

During the years following these publications the axiomatic treatment of subjective expected utility theory has been modified substantially through interactions between psychologists and economists, whereas the modification of Nash equilibrium has been left to economists for the most part. What explains this asymmetry? Our aim in this chapter is to offer some methodological reasons for the differences in the scholarly reception of the two main components of TGEb.

Among the first reactions to the formalization of decision and game theory were a number of (thought) experiments that sparked a debate over the normative and/or descriptive status of this research program. Classic examples include Allais's paradox, Ellsberg's paradox, and Schelling's focal points. Soon after that, an increasing number of laboratory experiments were conducted to systematically test the predictive accuracy of subjective expected utility theory, i.e., the individual-decision-theory side of the book. Several cognitive psychologists, in particular Lichtenstein and Slovic (1971), and Kahneman and Tversky (1971; 1979) reported the first critical results. Economists took these results of laboratory and field experiments seriously, and started to design experiments to test the robustness of the psychologists' findings (see e.g., Grether and Plott 1979). Hence, an interdisciplinary community of economists and psychologists proved pivotal to the rise of behavioral decision theory during the 1970s and 1980s.

On the other hand, the strategic side of the book has not prompted the same pattern of interdisciplinary exchange between economics and psychology. Although it has influenced many other disciplines in different ways, game theory has not been modified by psychology, as decision theory was. As a consequence, exchanges between economics and psychology on the basis of the common use of game theory remain limited to this day.

This is surprising, because in many ways game theory and expected utility theory share major features—the mathematical proofs of existence and uniqueness are foundational for both. In fact, expected utility theory was originally devised by von Neumann and Morgenstern to solve *games* with strategic uncertainty.

Our aim in this chapter is to consider the features of game-theoretic models that prevented them from being fully informed by psychological research, relative to decision-theoretic models. We argue that these features concern the belief concept embedded in the practice of equilibrium analysis that is specific to economics. In particular, we argue that the asymmetry in the reception of TGEb is attributable in part to the notion of belief implied in the Nash equilibrium.

We proceed as follows. Section 2 discusses and specifies the asymmetries in more detail. Section 3 presents two emblematic episodes that illustrate how the concept of equilibrium does not pertain to the explanatory toolbox of psychologists, whereas it is essential for economists. Section 4 identifies beliefs as a key construct that appears in both expected utility theory and game theory, but is understood differently: we argue that this mismatch made knowledge exchange between economists and psychologists more difficult. Section 5 explicitly analyzes expected utility theory and game theory as *boundary objects*—means of interdisciplinary exchange—and

clarifies their asymmetries by drawing on the previous sections. Section 6 briefly concludes the chapter.

## 2 Asymmetries

This section concerns the asymmetries between behavioral game theory and behavioral decision theory, with respect to their development as research programs and their impact in the scientific community. We focus first on the debate concerning the normative/descriptive status of game theory (GT), arguing that even if it evolved differently than expected utility theory (EUT), the interpretation of GT as a descriptive theory has a legitimate theoretical status and practical relevance, as the development of behavioral game theory shows. This point rules out an alternative hypothesis to explain the asymmetric developments of behavioral decision theory and behavioral game theory, namely that EUT allows straightforward empirical interpretation whereas GT is a normative or prescriptive theory, whose empirical status is complex and contested.

We then substantiate the asymmetry between the two fields with reference to the different role of psychologists in the genesis of experimental work. In doing so, we reject another possible explanation of the asymmetry, which suggests that psychologists did not find a fertile ground for collaboration in GT because the experimental results could all be accommodated within a framework of pure rationality. Our ultimate goal in this section is to demonstrate the need for a fundamental explanation of the difference between the two fields, and we argue that this is to be found in the different concepts of beliefs they adopt.

### 2.1 *Indications of Asymmetry and Alternative Explanations*

Some initial empirical support for the asymmetry we identify is provided in a recent paper by Doehne and Herfeld (2018) in which co-citation network analysis is used to study the diffusion of scientific innovations introduced by TGEb. The authors show how the spread of TGEb was mediated by scientific publications acting as “translators” to other fields, defining translators as publications that were pivotal in showing the relevance and potential of the innovations for a particular domain, sometimes creating a new field of inquiry as a result.

Doehne and Herfeld’s analysis clearly shows that the diffusion of rational choice theory to psychological fields was mediated by psychologists as translators. Interestingly, however, all the fields these translators created or impacted concern expected utility theory (behavioral decision theory, statistical decision theory, and mathematical psychology), but not game theory. There is, in fact, no psychological translator

in the domains of cooperative game theory, non-cooperative game theory, or theories of conflict and cooperation.

This asymmetry in how decision theory and game theory have spread is puzzling if one considers that a similar pattern seems to characterize the way in which researchers initially received EUT and GT: both were interpreted as descriptive theories yielding predictions that could be tested experimentally (see below). Indeed, in the opinion of many, the empirical results were problematic for both. Specifically, preference reversals, framing effects, fair divisions in the ultimatum game, and cooperation in the public goods game were challenging outcomes for the status of both EUT and GT as predictive and empirical theories.

A common reaction at this point is to say that GT has not promoted interaction between economists and psychologists, because GT should not be taken as a descriptive theory. However, such an explanation downplays some significant aspects of the research program that GT has prompted. On the one hand, the descriptive side of GT is flourishing, as the development of behavioral game theory indeed shows. If the reason why psychologists have not collaborated with game theorists is that GT is a normative theory, then we should not observe the development of behavioral game theory, either. On the other hand, behavioral game theory is a growing research area that has developed without a decisive contribution from psychology. Finally, many of the reasons why a normative reading would be more legitimate than its descriptive counterpart are discussed in the literature without leading to the conclusion that the normative interpretation of GT is the only legitimate one (on this point, see Guala 2006).

Another possible reaction is that the more complicated or contested empirical status of GT compared to EUT explains the asymmetry. The debate concerning the interpretation of empirical results in GT has indeed been more complex and contentious than that concerning EUT. One of the main reasons for this is that GT has a richer structure: it involves more than one individual decision maker, and therefore requires specification of the number of players, their interrelated strategies, related payoffs, and the information they have. This makes the problem of under-determination more severe: when the empirical results diverge from theoretically derived predictions, there are more potential culprits.

Although empirical tests of EUT also have to deal with the problem of under-determination (whether it is risk, loss, or regret that people avoid; whether it is beliefs rather than preferences), at least experimenters do not have to consider the options that derive from the richer structure of GT, as described above. In particular, they do not have to consider the utility involving beliefs about others: decision problems concerning lotteries do not involve other players. In short, tests of GT suffer more severely from under-determination issues than tests of EUT do.

The complex structure of GT further challenges its empirical status, as shown in the problem concerning the refinement of payoffs, or of what to include in *self-interest*. By way of an illustration, let us suppose that the participants in a laboratory one-shot Prisoner's Dilemma game decide to cooperate, contrary to the theoretical prediction of mutual defection. In this case, the payoffs could be refined to include the broader self-interest of the participants, such as preferences for satisfying the



other's preferences or perceived expectations, so that mutual cooperation is still explained as a result of rational play. The underlying justification is that this enables the experimenter to incorporate non-monetary outcomes that people care about. The worry, however, is that this strategy might make it impossible to test the theoretical predictions of rational play as opposed to selfish or narrowly-self-interested play. Not only would this make the theory unfalsifiable, it would also imply that the entire research program on behavioral game theory is misconceived, if not flawed.

However, concerns that game theory is empirically vacuous can be addressed (see e.g., Guala 2006). The refinement of people's utility function can be empirically justified to the extent that it captures more or less stable *patterns* of choices. In other words, it is possible to impose some empirical discipline on *ad hoc* postulations that any observed behavior is self-interested or maximizes utility in one way or another. In this respect, the rationales underlying experimental work and modifications in GT and in EUT are no different. In both cases, experimental analysis is supposed to test the choice models of individuals and, if necessary, provide indications of how to formulate more accurate descriptive models. Models of social preferences in behavioral game theory have been empirically evaluated in exactly this way (Camerer, 2003).

Our first conclusion is thus that the asymmetric developments of behavioral decision theory and behavioral game theory are attributable neither to the distinctly normative or prescriptive character of game theory's nor to its complex or contested empirical character compared to expected utility theory. Although the debates on EUT and GT proceeded independently, the goals of experimental work were similarly conceived. The main point in both cases is to test the predictions and, in case of divergence, suggest how to modify the theory so as to accommodate them in a systematic way.

## ***2.2 The Genesis of Experimental Work on Game Theory and Decision Theory, and how the Roles of Psychologists Compare***

Let us now turn to the genesis of experimental work. There were differences in the way in which experiments were introduced in EUT and GT—and in the responses they provoked. First, the body of experimental work on EUT that led to the development of behavioral economics was, in the main, instigated by psychologists. Economists took up the challenges raised in the empirical work almost immediately, and fruitful exchanges took place between the two disciplines.<sup>1</sup> The most influential theory to date remains prospect theory, which was developed by two psychologists.

In the case of GT, however, psychologists did not have as influential a role as they had in promoting experiments on EUT. Two mathematicians, Merrill Flood and Melvin Dresher, conducted the first experiments on the prisoner's dilemma in 1950. Another milestone paradigm in experimental game theory, the ultimatum game,

---

<sup>1</sup>See Lisciandra (2018) for an analysis of experiments on social preferences.

was developed by three economists—Werner Güth, Rolf Schmitterberger and Bernd Schwarze—in the context of testing Ariel Rubinstein’s model of bargaining.<sup>2</sup> After this, the ultimatum game provided a testbed for Fehr and Schmidt’s iniquity-aversion model in *A Theory of Fairness, Competition, and Cooperation* (1999), which has become extremely influential among behavioral and experimental economists.

A different story holds for social dilemma and public good games, which indeed have attracted social psychologists, sociologists and experimental economists, generating a lively multidisciplinary field (for a review, see Ledyard 1995). Here in fact, we find a pattern that closely resembles what happened in the experimental testing of EUT.

The interest of the economists—in particular of Mark Isaac and James Walker—in public good games was initially sparked by the experimental findings of a group of psychologists, under the guidance of Robin Dawes, and a group of sociologists including Gerald Marwell and Ruth Ames. Both groups were working on the so-called social dilemmas, which are a special version of public goods games (Dawes et al., 1977).<sup>3</sup>

The surprising observation among economists was that social dilemmas seemed not to suffer from the problem of free-riding behavior, as game theory predicts. Trying to prove them wrong, Isaac and Walker ended up with mixed results: their work did not reverse previous findings, but it did not confirm them, either. One major innovation that these economists brought to the table was the repeated game design, which turned out to influence individual contributions in experimental settings in interesting ways. The experiments indicated that, at least under certain conditions, free-riding behavior emerges with repetition (Isaac et al., 1985; Kim and Walker, 1984) in a way that is consistent with the theoretical expectations (coupled with an assumption that people initially make mistakes and eventually learn to play rationally).

Nevertheless, the results from public goods games were less conclusive than was desirable (see Ledyard 1995), and the overall picture that emerged from the empirical work was rather challenging for game theory. Among the main puzzling aspects, there was the issue of how to conceptualize fairness considerations or focal points in non-cooperative games, and how to explain the solutions to coordination games that have multiple Nash Equilibria, such as the Hi-Lo game.<sup>4</sup>

In a similar way as with EUT, these results seem to show that a purely game-theoretic approach that focuses exclusively on the rationality criteria is not well

<sup>2</sup>Daniel Kahneman was very “crestfallen” (Kahneman, 2014) when he got to know that he and his economists colleagues had been scooped by Güth et al. Kahneman published the scooped study as Kahneman et al. (1986).

<sup>3</sup>More generally, research on non-cooperative game theory—most importantly on Prisoner’s Dilemma and social dilemma—formed the field of conflict resolution. The field is problem-oriented and less theoretically integrated; that is, it involves researchers from multiple disciplines such as economics, psychology, political science, law and management (see Mnookin et al. 1995), but there has not been significant theoretical integration of insights from these disciplines.

<sup>4</sup>The Hi-Lo game is a game with two Nash equilibria, one of which has higher payoffs for both players than the other. Game theory, however, treats the two solutions as equally plausible, because strictly speaking they are both Nash equilibria. See Bacharach et al. (2006).

equipped to explain the phenomena observed in experiments. It may be that a more careful look at the process of beliefs formation could help to solve the foundational difficulties facing game theory (see Colman 2003).

The question thus remains: why have psychologists not engaged more critically with rational models of game theory in a way that would lead to theoretical modifications? After all, beliefs and desires lend themselves to psychological analysis, as the case of EUT shows.<sup>5</sup> Moreover, given that beliefs are at least as crucial in GT as in EUT, it is even more surprising not to observe the same pattern of exchange in the former. As we have argued above, the normative versus the descriptive interpretations cannot explain the divide between EUT and GT, hence the need for a different, more fundamental explanation of the asymmetry.

Below we will argue that one of the reasons for this outcome is that the very notion of beliefs is interpreted and used differently in the two theories. In the case of GT, desires are captured in the payoff of outcomes corresponding to strategy profiles, or different combinations of players' actions. When players choose actions such that their payoffs are maximized, the game is said to be "solved" because the system is in an equilibrium state—no unilateral deviation from that state yields benefit: the *beliefs* of each players about the other player's beliefs and actions are indeed correct. However, in many cases the analysis is silent about the way in which these beliefs have been reached: it finds a resting point, but it leaves unanswered the question of whether and how such a state can be reached by real players. Equilibrium analysis thus understood is a very familiar exercise to economists, but it is not equally prominent in the scientific practice of psychologists. If anything, psychologists are interested in *how* the equilibrium is reached, rather than in its theoretical foundations in strategic interactions *per se*. As we show in the following sections, this difference might have determined the divide between the two groups, and it could explain the asymmetries we have identified above.

### 3 Equilibrium Analysis: Two Illustrations

As we briefly noted at the end of the previous section, a crucial difference between the belief concept in GT and its counterpart in EUT is that the former is deeply connected to the equilibrium analysis of interactive play, whereas the latter is not. To understand this difference, we need first to characterize equilibrium analysis in economics, and then to illustrate how it figures in explanatory practices.

Equilibrium analysis, in its most abstract sense, is an approach to studying a given system's emergent properties as a result of interactions between its components. A market consisting of consumers and producers, and the markets comprising individ-

---

<sup>5</sup>It has indeed been argued that the "journey" from EUT to psychology and back was facilitated by the familiarity of psychologists with such concepts (Nagatsu and Małecka, 2019).

ual markets, have been the the main systems of interest in economics.<sup>6</sup> Game theory has extended economic analysis to any systems involving two or more interactive agents, and has shifted the focus from dynamic optimization to the mutual consistency of the system's components (Giocoli, 2003). As this shift implies, equilibrium analysis as applied by economists has changed in character during the history of economics.<sup>7</sup> Nevertheless, it is possible to characterize economic equilibrium analysis as an approach to studying the behavior of a system (or a set of systems) primarily in relation to its well-defined stable states. Our claim is not that only economists engage in this mode of analysis (in fact psychologists engage similarly in the study of bounded rationality), but that economists' way of doing so in game theory has created a specific disciplinary barrier in changing the nature of the belief concept in a non-transparent way.

To illustrate this point, we present two episodes contrasting responses of economists and psychologists to equilibrium analysis. The first one comes from game theory and the second one from experimental economics. We show how psychologists are reluctant to engage with that kind of explanations, then we compare the logic behind equilibrium-based explanations in game theory with optimization in decision theory. Let us start with the theoretical case.

### 3.1 *The By-Stander Effect*

The first case concerns the so-called by-stander effect in social psychology. In a popular introductory textbook on game theory, Martin Osborne (2004) builds a model of "reporting a crime," in which a group of homogeneous people decide whether or not to report a crime they have observed. The model serves as an illustration of how to use mixed-strategy equilibrium to solve coordination games with conflict (Sect. 4.8 in Osborne, 2004).

The game is played by  $n$  players, whose action set is {Call, Don't call}, and whose preference ordering over three outcomes is as follows: someone else calls  $\succ$  she calls  $\succ$  no one calls; mapping to expected utilities  $v > v - c > 0$  respectively, where  $v$  is the value she attaches to the crime being reported, and  $c$  is the cost she incurs to call herself, and  $v > c > 0$ . The model is intended to explain the brutal murder of Catherine ("Kitty") Genovese over a period of half an hour in New York City in 1964, discussed in the textbook.

---

<sup>6</sup>Studying optimization of consumers and firms as a result of profit and utility maximization under budget constraints is a type of equilibrium analysis, in which the systems in question are individual agents. In this chapter, however, we focus on the equilibrium analysis involving multiple agents.

<sup>7</sup>Hands (2010) argues that, independently from game theory, in the mid 20th century a significant change in the character of equilibrium analysis has occurred in consumer choice theory and general equilibrium theory in economics: the shift from finding a stable point toward which a system moves through a certain path, to finding a rest point in an dynamically related system of differential or difference equations.

The puzzle to be explained is why none of the 38 people who witnessed the incident reported the crime to the police. Osborne’s answer is that, other things being equal, the mixed-strategy Nash equilibrium of the game implies that the higher the number of people who witness the incident, the more likely it is that no one will report it. Specifically, it is an implication of the equilibrium condition that each player must be indifferent between calling and not calling in equilibrium:

$$v - c = v * \Pr\{\text{at least one other person calls}\} + 0 * \Pr\{\text{no one else calls}\}$$

or

$$v - c = v * (1 - \Pr\{\text{no one else calls}\})$$

or

$$\Pr\{\text{no one else calls}\} = c/v$$

Let us denote the probability that each person calls as  $p$ . The probability that no one else calls is the probability that every one of the  $n - 1$  people does not call, i.e.,  $(1 - p)^{n-1}$ . The equilibrium condition is  $(1 - p)^{n-1} = c/v$ , or

$$p = 1 - (c/v)^{1/(n-1)}$$

This is the unique, symmetric mixed-strategy equilibrium of the game, in which the probability of each person calling is  $p$  ( $1 > p > 0$ ). Given that this probability decreases as  $n$  increases, it is clear that the probability of each person’s reporting decreases as the group becomes larger. The more subtle point—that the probability that *no one* will report increases as the group becomes larger—is shown by focusing on any player  $i$ .

$$\Pr\{\text{no one calls}\} = \Pr\{i \text{ does not call}\} \times \Pr\{\text{no one else calls}\}$$

Recall that  $\Pr\{\text{no one else calls}\} = c/v$ , which is independent of the group size  $n$ . Because  $\Pr\{i \text{ does not call}\}$  increases as  $n$  increases (this is just  $1 - p$  for  $i$ ), one could conclude that the probability that no one calls also increases as  $n$  increases. The crime was unreported in the Genovese case *because of*—not *despite*—the large number of witnesses. In other words, it is better to have fewer people around if we hope to be rescued!

Osborne contrasts his explanation to three others offered by social psychologists to explain similar experimental findings about by-stander effects. The first one concerns the diffusion of responsibility—the larger the group size, the smaller is the psychological cost of not helping; the second is called audience inhibition—the larger the group, the greater the potential embarrassment of a helper if the help turns out to be inappropriate; the third one is about social influence—the larger the group size, the more likely it is that witnesses will infer from the inaction of the others that help is, in fact, not appropriate. Osborne subsumes these explanations in his model as the group size ( $n$ ) either raising the expected cost ( $c$ ) or reducing the expected benefit ( $v$ ) of

helping.<sup>8</sup> He then points out that the implication of a mixed Nash equilibrium—that the larger the group size the less likely it is that at least one person will report a crime—holds even if the values of  $v$  and  $c$  are *independent* of the group size  $n$  or, equivalently, even if the three psychological effects of group size are absent.

Osborne's point, therefore, is not that these psychological explanations are wrong or redundant—some or all of them may well be contributing factors—but rather that they all miss the crucial notion of an *equilibrium*. Osborne thus concludes:

Whether any given person intervenes depends on the probability she assigns to some other person's intervening. In an equilibrium each person must be indifferent between intervening and not intervening, and as we have seen this condition leads inexorably to the conclusion that an increase in group size reduces the probability that at least one person intervenes. (pp. 133–134)

For the current purposes, it is not important whether Osborne's is the correct explanation of the by-stander effect in general, or of the Genovese case in particular. Nor is it important whether his account is more unifying than those offered by psychologists in some sense that philosophers of science specify. The moral of this illustration is rather that Osborne clearly contrasts equilibrium analysis that is central in economics to typical social-psychological explanations with no explicit equilibrium analysis. Note that the asymmetry does not imply that equilibrium analysis excludes any concepts of beliefs. On the contrary, Osborne states that the agent assigns a probability to an event in order to be indifferent between two actions. This point requires some mental ascription to the agent he is modeling. However, as we illustrate in the next section, these psychological concepts, beliefs in particular, are understood and used in economics very differently than in psychology. Before discussing that, we present another episode that highlights the specificity of equilibrium analysis.

### 3.2 *Almost Like Magic: The $N^*$ Game*

Another illustration of how equilibrium analysis is distant to psychologists comes from Nobel Laureate Daniel Kahneman, arguably the most authoritative psychologist to talk about economics. In his 2002 autobiography, he recalls the “magic” he observed in an experiment he conducted with economists Richard Thaler and James Brander. The experiment, called the  $N^*$  game, is an  $N$ -player symmetric market-entry coordination game without communication: the market is profitable for entrants, but the marginal profit from entry decreases as the number of entrants increases, and beyond a certain market capacity (denoted by  $N^*$ ) profit becomes negative. In the original experiment,  $N = 15$  and  $N^*$  was changed over a period of repetition within the range of  $12 > N^* > 3$ ; the payoff to each person was \$.25 if one did not enter the market, and  $[\$.25 + .5(N^* - E)]$  if one did, where  $E$  is the

---

<sup>8</sup>In fact, the third explanation can be cast in a game theoretic model of pluralistic ignorance, but we will follow Osborne's presentation here.

number of total entrants.<sup>9</sup> If  $E = N^*$ —if the number of actual entrants equals the capacity—both entrants and non-entrants receives the same \$.25 payoff. To Kahneman’s great surprise,  $E$  quickly converged to  $N^*$  in a few rounds, and stayed within the range  $N^* - 2 < E < N^* + 2$  in the vast majority of trials. Of course, this is the implication of a mixed-strategy Nash equilibrium: each player decides to enter with a probability such that the expected payoff from entry equals that from non-entry—and is indifferent between entering and not entering in a steady state. The aggregate outcome from such individual strategies will result in  $E \approx N^*$ .

In Kahneman’s words, “[o]bserving the regularity of behavior in these markets was a bewildering experience—to a psychologist, it looked almost like magic.” (p. 12 Kahneman 1988) And “it took me some time to realize that the magic we were observing was an equilibrium: the pattern we saw existed because no other pattern could be sustained.” (2014) Moreover, the debriefing conversations revealed that most participants’ accounts of their own “winning” strategies were unfounded and had no clear connection with the equilibrium results. In other words, “[t]he equilibrium outcome (which would be generated by the optimal policies of rational players) was produced in this case by a group of excited and confused people, who simply did not know what they were doing.” (Kahneman 1988, p. 12)

Kahneman summarizes his lessons from this study as follows:

Psychologists are trained to believe that aggregate phenomena can be explained by finding some relevant regularity in individual behavior. The  $N^*$  game provided me with first-hand experience of a clear failure of this belief. The only solid explanation of the results of the  $N^*$  game belongs to a type that is quite familiar to economists, but not to other social scientists.[...] The cognitive psychologist discovers that he has essentially nothing of interest to contribute, and that his bag of intellectual tools lack the powerful instrument of equilibrium explanations. (Kahneman 1988, pp.12–13)

This is a surprisingly unguarded remark, coming from the cognitive psychologist who has so forcefully and successfully challenged EUT, or the vNM utility notion used in the derivation of the mixed-strategy Nash equilibrium. One could argue that the results are also problematic for economists because they cannot explain the gap between the equilibrium results and the (confused) self-reports of the participants of the  $N^*$  game, either. However, since economists typically neither demand players’ conscious awareness of their reasoning processes, nor value self-reports as reliable evidence, this gap seems less problematic for them. In any case, equilibrium analysis yields accurate prediction, whereas no psychological account alone fills the gap.<sup>10</sup>

As in the Osborne case, the fact that equilibrium analysis explains some phenomena, whereas psychological accounts do not does not imply that the former involves no psychological concepts. As we will show in the next section, the concept of beliefs is implicit in the use of the mixed-strategy Nash equilibrium.

<sup>9</sup>We used the numbers as in Kahneman (1988), which are slightly different from those presented in his Nobel autobiography.

<sup>10</sup>For modern discussion on this game, see Dhami (2016). Interestingly, no other behavioral economics textbooks than this advanced one discusses  $N^*$  game.

### 3.3 *The Methodological Rationale Behind Equilibrium Analysis*

The two episodes described above encourage the use of equilibrium analysis in showing how successful it can be as an explanation or prediction. This is not the only reason behind the use of equilibrium-based accounts, which do not always succeed in predicting or explaining the phenomena of interest, be they market aggregates or collective behaviors. Economists therefore resort to a more general, methodological rationale. According to Herbert Gintis, for example, economists explain aggregate phenomena in terms of an equilibrium “not because it accurately reflects actual economic conditions, but rather because it is instructive to understand when it does not, and why” (p. 251 Gintis 2017). In other words, equilibrium analysis is justified as providing an empirical benchmark and a heuristic for explaining deviations when they exist.

In fact, this methodological rationale is very similar to the rationale behind the use of *optimization* as a benchmark in behavioral decision theory, in which a clear prediction from the canonical optimization model (e.g. EUT or exponential time-discounting) is derived, then a deviation is experimentally demonstrated, and finally explanations in terms of individual psychological biases (e.g., loss aversion or present bias) are provided. It is also possible to extend the same “benchmark-deviation-biases” strategy to explaining aggregate-level phenomena. In their famous mug cup study, for example, Kahneman et al. (1990) first derived a clear prediction from market-equilibrium analysis: because mug cups are *randomly* given to half of a group of the participants, there is a 50-50 chance that the new owners will value their cups more than those who did not receive one. Therefore, half of the mug cups would be voluntarily traded (benchmark). The researchers then demonstrated that the volume of the trade was significantly less than half (deviation), which they finally explained in terms of loss aversion or anticipated regret making the cup owners more reluctant to sell relative to the willingness to pay of potential buyers (bias). The endowment effect thus constructed is an aggregate-level phenomenon, but it is explained in terms of individual bias.

The upshot of this section is that, although equilibrium analysis is a very economic way of explaining aggregate phenomena, compared to social and cognitive-psychological approaches, its underlying methodological rationale (benchmark and heuristic) is no different from that of psychologists engaged in behavioral decision research. In what sense, then, are these two illustrations of game-theoretic equilibrium analysis alien to psychologists? As we hinted in this section, we now argue that the conceptual difference between beliefs in equilibrium and beliefs in individual decision-making may have been the main barrier to productive collaboration (in a broad sense) between economists and psychologists in behavioral game theory.



## 4 Beliefs in Equilibrium and Beliefs in Individual Decision Making

In the previous section we presented two case histories, one of a formal model (told by an economist) and the other of an experiment (told by a psychologist), both of which highlight equilibrium analysis as a distinctively economics-based explanatory style of observed aggregate patterns of human behavior. However, this does not mean that equilibrium analysis is void of psychological constructs. On the contrary, finding an equilibrium solution to a given game necessitates conceptualization of beliefs as a theoretical construct, as we will show. Our point is rather that such a belief concept is distinct from its psychological counterpart that features in EUT. Our aim in this section is to make this conceptual difference explicit. First we explain how the belief concept is defined in game theory, and how it is implied in the way equilibrium analysis is conducted, then we contrast it to its counterpart in EUT.

Hal Varian (2010), in his popular intermediate-level microeconomics textbook, informally defines a Nash equilibrium in a two-player strategic form game as follows:

a pair of strategies is a **Nash equilibrium** if A's choice is optimal, given B's choice, and B's choice is optimal given A's choice. (p. 524)

Here is Osborne's (2004) slightly more technical and precise definition in the introductory textbook on game theory discussed above:

A *Nash equilibrium* is an action profile  $a^*$  with the property that no player  $i$  can do better by choosing an action different from  $a_i^*$ , given that every other player  $j$  adheres to  $a_j^*$ . (p. 22)

Although beliefs do not feature in these definitions, Varian (2010) makes a significant observation immediately following his definition quoted above.

Remember that neither person knows what the other person will do when he has to make his own choice of strategy. But each person may have some expectation about what the other person's choice will be. A Nash equilibrium can be interpreted as a pair of expectations about each person's choice such that, when the other person's choice is revealed, neither individual wants to change his behavior. (pp. 524–5)

In other words, players do not know what the other person will do, but in equilibrium they act *as if* they did: their beliefs about others' actions are coordinated in the sense that they are the same and true in equilibrium. In fact, coordinated beliefs are the second component in the notion of Nash equilibrium, as Osborne (2004, p.20) explicitly mentions, and they are made even more explicit in Varian's (1992) advanced microeconomics textbook (p. 265):

A **Nash equilibrium** consists of probability beliefs  $(\pi_r, \pi_c)$  over strategies, and probability of choosing strategies  $(p_r, p_c)$ , such that:

1. the beliefs are correct:  $p_r = \pi_r$  and  $p_c = \pi_c$  for all  $r$  and  $c$ ; and,
2. each player is choosing  $(p_r)$  and  $(p_c)$  so as to maximize his expected utility given his beliefs.

Varian refers to two players, Row and Column;  $p_r$  denotes the probability of Row playing  $r$ ; and  $\pi_c$  denotes Row's subjective probability distribution over Column's choices, i.e. Row's beliefs about Column's behavior, and similarly for Column. Of note here is that this definition requires each player's subjective beliefs about others' choices to coincide with their actual choices.

Varian warns the advanced reader that a more conventional definition of a Nash equilibrium—such as his in the intermediate textbook and Osborne's quoted above—is misleading “since the distinction between the beliefs of the agents and the actions of the agents is blurred” (1992, p. 266). What, then, is the nature of these beliefs thus distinguished from the actions of agents? Do agents really “know” other agents' actions in some philosophically or psychologically well-founded sense?

These questions concern the interpretation of a Nash equilibrium and are not part of any formal definition. In fact, during the 1980s there was a debate between psychologists and game theorists concerning the exact interpretation of beliefs in game theory. In brief, psychologists such as Kadane and Larkey (1982) criticized the assumption that players' beliefs about each others' beliefs and behavior in equilibrium were necessarily true, whereas game theorists such as Harsanyi defended this interpretation of beliefs as necessary theoretical apparatus to derive a solution to any given game (see Grüne-Yanoff and Lehtinen 2012; Morris 1995 for a detailed discussion of the reasons why economists adopt the common prior assumption).

In contrast to this philosophical and theoretical literature on the nature and epistemological foundations of beliefs in game theory, the pronouncements of many practicing economists on these issues are neither explicit nor eloquent. In fact, there is some indication that economists are not primarily concerned with such conceptual and epistemological questions. Instead, it seems that the specification of a Nash equilibrium is primarily driven by the need to find a solution to a game that is “in some sense, in equilibrium” (p.264 Varian 1992), or that satisfies a “natural consistency requirement” (p. 265). Hence, in an idealized setting a Nash equilibrium therefore corresponds to a “*steady state*” (p.20 Osborne 2004) in which there is no pressure for change. This, of course, is a familiar way of modeling economic phenomena for economists. Students of microeconomics typically learn about the theory of the market before learning about game theory: they study the concept of equilibrium by deriving it from the consumers' demand function and the firms' supply function, without conducting a thorough analysis of the epistemological foundations of belief formation, and only *then* do they study the concept of a Nash equilibrium in the context of game theory and as a generalization of the Cournot equilibrium.<sup>11</sup> Thus, the derivation of a Nash equilibrium is a “natural” extension of how economists model aggregate outcomes as equilibria, or steady states.

This explanatory practice of economists distinguishes the notions of beliefs in the Nash equilibrium and in EUT. In the latter they are represented as a distribution of subjective probabilities over the state of the world, which follow certain rational

<sup>11</sup> Although there is an alternative way of organizing microeconomics, proceeding from the optimizing individual to strategic interactions to market interactions (e.g. Bauman and Klein 2010), such an organization seems still minority.

requirements such as Bayesian updating and basic probability calculus. Subjective probability may even be weighted over objective probability, as in Tversky and Kahneman's Cumulative Prospect Theory (1992), because of the psychological principle of decreasing marginal sensitivity from certainty as a reference point, for example. EUT is a useful theory that allows psychologists to dovetail their insights because the concept of beliefs it espouses is a natural extension of the common-sense understanding with a formal representation, and psychologists are used to dealing with it. In contrast, beliefs in game theory—more specifically in a Nash equilibrium—do not correspond strongly with the intuitive understanding of beliefs: they have been derived from a discipline-specific drive to identify the properties of a system in (and out of) a steady state.

This drive is at the heart of the assumption that aggregate-level interactions among purposeful agents will, in the idealized condition, settle in some steady state. As a result, the implied notion of beliefs is not easily commensurable with the interpretations of psychologists, according to which they are either personal priors (the subjectivist view) or are based on some subject-independent features of the external world (the frequentist view). This explains why psychologists are typically unwilling to accept the legitimacy, let alone the usefulness, of a concept of beliefs that is so different from the concept they commonly adopt.

We do not mean to imply that game theorists are conceptually sloppy about what they mean by beliefs when we refer to the concept as ambiguous. On the contrary, beliefs are variably but precisely defined in Bayesian games, extensive games, and so on. The point is that the foundations of beliefs in game theory are primarily based on the practice of equilibrium analysis in economics in general, and the Nash equilibrium in game theory in particular. Within this paradigm, there are alternative ways of conceptualizing beliefs. For example, although the rationalizability approach “assumes that the players know each others’ preferences, and considers what each player can deduce about the other players’ actions from their rationality and their knowledge of each other’s rationality” (p.21 Osborne 2004), Osborne clearly considers it optional and an alternative for economists. Economists, however, simply use a Nash equilibrium to model a steady state that will be reached through the interactions of experienced players. Ultimately, it is regarded as “a matter of judgment” (p.24 Osborne 2004) whether or not the notion is appropriate to model a given situation, not as a matter of whether or not epistemic foundations can be found, or the implied concept of beliefs can be reconciled with other existing belief concepts such as the subjectivist or the frequentist.

## 5 Game Theory: A Computational Template as a Boundary Object

In this section, we apply a *computational template* to analyze the asymmetries between the ways EUT and GT functioned to facilitate interdisciplinary collaboration between economists and psychologists. Our observation is that focusing on how

TGEB as an innovative monograph was “diffused” across the disciplines (Doehne and Herfeld, 2018) cannot capture the asymmetries in which we are interested. First of all, we need a finer-grained unit of analysis than a publication of TGEB that includes both EUT and GT.

Second, our interest is not in the way in which TGEB was transferred from one domain to another, but in the conditions that allowed a community of researchers, including economists and psychologists, to work together and contribute to the shaping of the field of behavioral decision theory, and conversely in the factors that hindered the same research community to collaborate similarly to develop behavioral game theory. For these reasons, we think it is more fruitful to focus on a smaller unit of analysis (the computational template) and its function of *mediating* interdisciplinary collaboration rather than treating it as an object that was developed in one field and then was transferred to other domains. Let us start with the notion of templates.

Paul Humphreys (2004) was the first to use the notion of templates—in contrast to a theory and a model—as a unit of analysis to study progress in the physical sciences and their real-world applications. He notes that much progress in these fields is driven by the invention and deployment of tractable mathematical formulae, which he calls *computational templates*. A computational template is to be distinguished from a *theoretical* template in that it is a representational device limited to a mathematical or computational interpretation, whereas the theoretical template is interpreted within the domain of a theory (Humphreys 2018).

In addition to being mathematically tractable, computational templates “can be considered from a purely syntactic perspective” (p.59 Humphreys 2004), allowing some “flexibility and degree of independence from subject matter” (p. 67). Knuutila and Loettgers (2014) identify this syntactic nature of computational templates as a facilitator of their interdisciplinary transfer. Grüne-Yanoff (2011) similarly analyzes evolutionary game theory as an evolving template traveling from economics to biology, and back. Within this literature, interdisciplinary exchange is analyzed in terms of traveling computational templates, facilitated by their syntactic, subject-neutral nature.

We follow this line of thinking in focusing on computational templates as a unit of analysis. However, rather than seeing them as traveling syntactic objects, we find it more relevant here to construe them as objects that *mediate* coordination and collaboration between different scientific communities despite existing epistemic and conceptual tensions. This notion of “mediating” units goes back to the idea of *boundary objects* in the sociology of science, introduced by Star and Griesemer (1989) and further developed by researchers such as Wenger (1998). According to Star and Griesemer (1989):

“Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use.” (p. 393)

What is curious about GT and EUT is that *prima facie* they share the features that would make them both boundary objects and computational templates: given that the variables they adopt, i.e., beliefs, preferences, and choice, are the same, they should be equally applicable across the same domains. At the very least, if success is achieved in the case of EUT and psychology, the same should apply to the case of GT and psychology. Moreover, with respect to the formalism they employ, they also seem to share a similar mathematical machinery, which would suggest that they should be equally suitable as means of interdisciplinary transfer.

However, we maintain that EUT works well as a boundary object and a computational template in economics and psychology, but that the same does not apply to GT. Our explanation of asymmetric collaboration can be summarized thus: as a boundary object, EUT facilitates collaboration between economists and psychologists because the *interpretation* of the psychological concepts employed in EUT (beliefs, preferences and choices) is largely shared between these two communities (as well as among philosophers and lay people). They differ in terms of practices such as experimental procedures, but nevertheless, psychologists were heavily involved in modifying EUT in the face of anomalous empirical findings. The level of abstraction was not an obstacle, but rather presented an opportunity for psychologists to contribute their expertise in modifying the model. Psychologists probably have less faith in individual optimization than economists have, but the ideal type of optimizing individuals was close enough to allow their involvement.<sup>12</sup>

Rather than focusing on different styles of modeling (economists' optimization-based vs. psychologists' process-based modeling), we highlight the centrality of equilibrium analysis as economic practice, and the implied notion of beliefs in equilibrium, which constitutes a conceptual obstacle hindering psychologists from participating in the modification of GT drawing on psychological expertise.

In this sense, EUT functioned as a boundary object that was flexible enough to be used by economists and psychologists with different theoretical and experimental backgrounds, while at the same time its substantive interpretation was robustly shared to allow for fruitful collaboration. On the other hand, the requirements on GT regarding use of the concept of beliefs, which derive from the formal requirements of equilibrium-based analysis, have no counterpart in psychology. Although GT as a computational template traveled across many domains, it did not function as a fruitful boundary object between psychology and economics: the belief concept was not robust enough to allow for overlapping interpretations.

---

<sup>12</sup>Of course, some psychologists refuse to collaborate with economists precisely because they reject optimization as even an ideal type. Our point is that EUT, in particular its belief concept, was close enough to that of psychologists that critical mass of psychologists collaborated.

## 6 Conclusions

The interdisciplinary collaboration between psychologists and economists mediated by EUT gave rise to behavioral decision theory, which in turn has formed a core part of behavioral economics. The empirical and theoretical work of cognitive psychologists was crucial in this development (Heukelom, 2014). The exchange took place as a modification of EUT as a boundary object. Game theory did not function in the same way, despite its common origin. In particular, the extent to which insights from psychology have modified GT is limited, although GT did travel to psychology in the form of experimental paradigms such as the Prisoner's Dilemma and other social dilemmas. As a consequence, behavioral game theory as a subfield of behavioral economics remains, for the most part, a game for economists.

We have explained this asymmetry as a result of the peculiarities of equilibrium analysis in the practice of economists, which embed the conceptual differences in how economists and psychologists understand beliefs—a seemingly unproblematic notion. If our explanation is on the right track, there are implications for the methodology of interdisciplinary sciences. Specifically, those answering a common call for a more interdisciplinary approach in the behavioral and social sciences—which we endorse and encourage—will need to pay more careful attention to the conceptual differences that may constitute a barrier to this admirable cause. Gintis (2017) identifies the reluctance of non-economists to fully embrace common-core analytical foundations—such as decision theory and game theory—as the main obstacles against integrating or unifying the behavioral sciences (meaning the social sciences plus sociobiology). However, this reluctance may not be solely attributable to disciplinary rent-seeking or a lack of mathematical training on the part of psychologists and other social scientists: conceptual differences may play a key role as a specific kind of what MacLeod (2016) refers to as *cognitive obstacles* to interdisciplinarity. Recognizing such a conceptual difference is a first step in designing more effective interdisciplinary methodological strategies.

## References

- Bacharach, M., N. Gold, and R. Sugden. 2006. *Beyond individual choice: Teams and frames in game theory*. Princeton, N.J.: Princeton University Press.
- Bauman, Y., and G. Klein. 2010. *The cartoon introduction to economics: Volume One: Microeconomics*. Hill and Wang.
- Camerer, C.F. 2003. *Behavioral game theory*. Princeton, NJ: Princeton University Press.
- Colman, A.M. 2003. Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences* 26: 139–+.
- Dawes, R.M., J. McTavish, and H. Shaklee. 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology* 35 (1): 1–11.
- Dhami, S. 2016. *The foundations of behavioral economic analysis*. Oxford: Oxford University Press.

- Doehne, M., and C. Herfeld. 2018. The diffusion of scientific innovations: A role typology. *Studies in History and Philosophy of Science Part A*.
- Fehr, E., and K.M. Schmidt. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114 (3): 817–868.
- Gintis, H. 2017. *Individuality and entanglement: The moral and material bases of social life*. Princeton, NJ: Princeton University Press.
- Giocoli, N. 2003. *Modeling rational agents: From interwar economics to early modern game theory*. Cheltenham, UK: Edward Elgar Publishing.
- Grether, D., and C.R. Plott. 1979. Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69: 623–638.
- Grüne-Yanoff, T. 2011. Models as Products of Interdisciplinary Exchange: Evidence From Evolutionary Game Theory. *Studies in History and Philosophy of Science Part A* 42 (2): 386–397.
- Grüne-Yanoff, T., and A. Lehtinen. 2012. Philosophy of game theory. *Philosophy of Economics* 531–576.
- Guala, F. 2006. Has game theory been refuted? *The Journal of Philosophy* 103(5): 239–263.
- Hands, D.W. 2010. Stabilizing consumer choice: the role of ‘true dynamic stability’ and related concepts in the history of consumer choice theory. *The European Journal of the History of Economic Thought* 17 (2): 313–343.
- Heukelom, F. 2014. *Behavioral economics: A history*. Cambridge: Cambridge University Press.
- Humphreys, P. 2004. *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press.
- Isaac, R.M., K.F. McCue, and C.R. Plott. 1985. Public goods provision in an experimental environment. *Journal of Public Economics* 26 (1): 51–74.
- Kadane, J.B., and P.D. Larkey. 1982. Subjective probability and the theory of games. *Management Science* 28 (2): 113–120.
- Kahneman, D. 1988. Experimental economics: A psychological perspective. In *Bounded Rational Behavior in Experimental Games and Markets*, ed. Tietz, R., W. Albers, and R. Seltenpages, 11–20. Berlin: Springer.
- Kahneman, D. 2014. Daniel Kahneman - Biographical. *Nobel Media AB*.
- Kahneman, D., J.L. Knetsch, and R.H. Thaler. 1986. Fairness and the assumptions of economics. *The Journal of Business* 59 (4): S285–S300.
- Kahneman, D., J.L. Knetsch, and R.H. Thaler. 1990. Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy* 98 (6): 1325–1348.
- Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2): 263–292.
- Kim, O., and M. Walker. 1984. The free rider problem: Experimental evidence. *Public Choice* 43 (1): 3–24.
- Knuutila, T., and A. Loettgers. 2014. Magnets, spins, and neurons: The dissemination of model templates across disciplines. *The Monist* 97 (3): 280–300.
- Ledyard, J. 1995. Public goods: A survey of experimental research. In *The handbook of experimental economics*, Kagel, J.H. and A.E. Roth, 111–194. Princeton University Press.
- Lichtenstein, S., and P. Slovic. 1971. Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89: 46–55.
- Lisciandra, C. 2018. The role of psychology in behavioral economics: The case of social preferences. *Studies in History and Philosophy of Science Part A* 72: 11–21.
- MacLeod, M. 2016. What makes interdisciplinarity difficult? some consequences of domain specificity in interdisciplinary practice. *Synthese*.
- Mnookin, R.H., L. Ross, K.J. Arrow, and A. Tversky (eds.). 1995. *Barriers to conflict resolution*. New York, NY: Norton.
- Morris, S. 1995. The common prior assumption in economic theory. *Economics and Philosophy* 11 (2): 227–253.

- Nagatsu, M., and M. Małecka. 2019. How behavioural research has informed consumer law: the many faces of behavioural research. In *Research Methods in Consumer Law: A Handbook, Handbooks of Research Methods in Law series, chapter 11*, ed. H.-W. Micklitz, A.-L. Sibony, and F. Esposito, 357–398. Cheltenham, UK: Edward Elgar Publishing.
- Osborne, M.J. 2004. *An introduction to game theory*. Oxford: Oxford University Press.
- Star, S.L., and J.R. Griesemer. 1989. Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907–39. *Social Studies of Science* 19 (3): 387–420.
- Tversky, A., and D. Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* 76 (2): 105–110.
- Tversky, A., and D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of utility. *Journal of Risk and Uncertainty* 5: 297–323.
- Varian, H.R. 1992. *Microeconomic analysis*, 3rd ed. New York: W.W. Norton & Co.
- Varian, H.R. 2010. *Intermediate microeconomics; A modern approach*, 8th ed. New York: W.W. Norton & Co.
- Von Neumann, J., and O. Morgenstern. 2004. *Theory of games and economic behavior*, 60th anniversary ed. Princeton, N.J.: Princeton University Press.
- Wenger, E. 1998. *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.



# *Homo Economicus* Under Multiple Pressures



Uskali Mäki

**Abstract** The issues around *homo economicus* are deep, and the debates over it have been heated. The chapter identifies four pressures that have shaped these concerns and conversations—deriving from economics itself, from the common sense, from disciplines other than economics (such as sociology, experimental psychology, and cognitive neuroscience), and from important features and processes in social reality (such as self-interest as a social norm, exposure to money, and the marketization of society). Economic theories assuming *homo economicus* find support from such trends in society, while the common sense and other disciplines largely speak against it. Among the themes examined are the applicability of the model of *homo economicus* across a negotiable range of phenomena, the negligibility of the falsity of its idealizing assumptions, and the consequences of designing policies on the presumption that people are sufficiently close approximations to *homo economicus*.

**Keywords** *Homo economicus* · Idealised model · Economic theory · Common sense · Psychology · Marketisation · Applicability · Negligibility

## 1 Introduction

*Homo economicus* has always been a contested creature, regardless of its particular incarnations. Some of the contestations are more or less constant, recurring across the history of the concept. Others are historically more specific, prompted by changes in economics itself or in neighbouring disciplines or in larger cultural circumstances. The same applies to the various justifications that have been proposed in support of

---

This paper derives from talks given at University of Tokyo, December 2016; Poznan, April, 2019; Lyon, December 2019; TINT, University of Helsinki, May 2020. Thanks for very helpful comments and suggestions on earlier drafts go to Wade Hands, Michiru Nagatsu, Emrah Aydinonat, Magdalena Malecka, and Pablo Jensen.

---

U. Mäki (✉)  
School of Philosophy, Nankai University, Tianjin, China  
e-mail: [uskali.maki@helsinki.fi](mailto:uskali.maki@helsinki.fi)

Practical Philosophy, University of Helsinki, Hämeenlinna, Finland

the concept. It is obvious that the story of *homo economicus* is also a story of the various tensions and debates that have accompanied its development.

That *homo economicus* is subject to ongoing debates is not surprising in the least—it would be surprising if this were not the case. The reason is also quite obvious: *homo economicus* appears to be about us—about human beings, their cognitive and moral properties, and these are also the subject not only for several scientific research fields, but also for ordinary commonsense conceptions of what we humans are like, and how we should be. On the other hand, *homo economicus* is not like us at all. This is rather easy for us to see, and this makes the concept so deeply puzzling.

Things are made complicated by variations and uncertainties in what precise concept of *homo economicus* is being discussed and what precise claims are supposed to be made about human behaviour when using the concept.

I will address these issues by framing them as deriving from “pressures” on the concept of *homo economicus*. I identify four such pressures: one from economics itself (what economics needs to assume to accomplish what it is expected to deliver); another from the common sense (with a rich nuanced image of our moral and cognitive qualities); a third from other disciplines (what they appear to imply about human behaviour); and yet another from aspects of social reality (such as the norm of self-interest, money, and marketization as constitutive of our culture). I will examine these pressures and point out affinities and tensions between them.

But first we need to have some initial clarity of what we are talking about and of what some of the issues are. The rough ideas presented next will be refined as we proceed to subsequent sections.

## 2 What Are We Talking About, and How?

So what are we talking about when we talk about *homo economicus*? From an abstract metaphysical point of view, we seem to be talking about a *thing*—at least apparently a possible human being—in terms of a set of *properties* attributed to that thing. These properties are described in terms of “assumptions” of a theory or model. This brings out an important qualification in how we should answer the question about what we are talking about here.

On the one hand, we seem to be directly talking about “models of man” as the phrase used to go. *Homo economicus* is a model, an imagined thing the (imagined) properties of which are characterized by the assumptions of the model. And as we know, this must be put in the plural: there are many models of *homo economicus*, depending on the schools of economic thought and the stage of their development (see e.g. Morgan 2006). On the other hand, when using these models, we seem to be indirectly talking about real things, such as people—or “agents” more generally—populating the world. We may ask whether this or that human individual is, or behaves in some aspect of his/her life as, a *homo economicus*; but by this we mean to ask whether this or that model of *homo economicus* has an appropriate relationship with

the properties of that individual such that we could characterise it (e.g., her or him) as a *homo economicus*.

I will primarily be talking about *homo economicus* as a model—or rather a set of models—of human (and possibly other) beings, and secondarily about those beings in the world being *homo economici* or being like *homo economicus*. The assumptions of such a model describe the model, and they play a key role when asking questions about real beings.

The set of such assumptions is seldom articulated in full, and they are seldom formulated in a way that captures much psychological nuance. The following list encapsulates an extended set of properties and puts them in a way that is recognizable to most people without an education in economics (such a list will therefore be useful for examining the relations between *homo economicus* and the common sense). *Homo economicus* is

- Determined: Knows what wants, with no conflicts in motives
- Informed: Knows how to get what wants
- Chooser: More options is better
- Self-seeker: Pursues own interests only
- Greedy: More is better
- Calculative: Likes to, and is able to, calculate
- Emotionless: No feelings of shame, fear, honour, empathy
- Amoral: Doesn't recognize ethical principles
- Asocial: Loner, blind to social norms, community values, common interests

Some of these characteristics are ordinarily found venerable, some others less so, even to the point of being reprehensible. Those educated in economics have learnt more streamlined formal versions of *homo economicus* that conceal much of the psychological nuance—as well as potentially morally dubious aspects—that are explicit in the above list. These versions include those that maximize their expected utility (or expected returns in the case of business firms), have “well-ordered preferences” (understood in terms such as completeness and transitivity), are informed about the available options and the probabilities of their consequences, and are capable of solving any given optimization problem.

For framing and settling controversies over *homo economicus*, there is a prior issue of how to interpret its contents regarding whether it is at least partly about internal mental features and processes in human agents, or rather about their overt behaviour. *Internalist* versions read it as representing internal cognitive capacities, dispositions and processes that are being invoked when using *homo economicus* for whatever purposes, such as explanation, prediction, design or what have you. *Externalist* versions read *homo economicus* as being quiet about anything mental, and instead as representing the structure of agents' manifest behaviour. Curiously, and somewhat misleadingly, given the vocabulary that is commonly used when discussing these issues, internalist perspectives seem fit for “behavioural” approaches, while externalist perspectives are favoured by “behaviourist” approaches. The difference between the two will play a role for the power of pressures on *homo economicus* as we will see in Sect. 3.

When putting forth assumptions about *homo economicus*, what exactly is being said? And what exactly is being asserted about human behaviour? These questions can be asked and answered from several perspectives. One deals with *quantification*. How many? All, many, few, none, which? This can concern *cultures* within which people live and act, such as more collectivist and more individualist cultures. Does *homo economicus* apply to all cultures, or just to some, or perhaps none? It can concern *times*—past, present and future. Does *homo economicus* apply always, or just to limited time periods, or perhaps never? It can concern *domains of action*, such as market and non-market spheres of society. Does *homo economicus* apply to all spheres, to just some, or none? Finally, the issue can concern *individuals*. Does *homo economicus* apply to all individuals, or just to many or some, or perhaps none? Does it apply to groups, representative agents, institutions—and if so, which?

The perspective from quantification must be supplemented by another one. From this other perspective, we may refine the contents of the expression I used in the previous paragraph when asking questions about quantification in “applying” the notion of *homo economicus*. “Apply” and its derivatives such as “applicable” are indeed far from unambiguous concepts. What exactly is it for *homo economicus* to apply or for it to be applicable? What is being said and asserted when applying the notion, and how does this relate to the real world, whatever the quantification?

On the face of it, it seems obvious that the assumptions of *homo economicus* are *idealizations*, and so they appear false, violating facts of the matter about human behaviour. Our preferences are not complete or perfectly transitive. We are not properly informed about the relevant probabilities, nor are most of our information processing capacities terribly impressive. The power of emotions, moral sentiments and pro-social preferences in moving us is often much greater than nil. And so on.

While the contents of *homo economicus* is put in terms of such idealizations, it is evident that there is no intended assertion that real people are like that, such that the idealizations would be claimed to be exactly true about them. On the contrary, they are generally recognized as false if interpreted as truth-valued statements about real people. Users of the model may also entertain other options. One is to suggest that the model of *homo economicus* is not used and should not be used for making any claims at all about people’s behaviour; the model is perhaps considered just a “fictional” component in a larger theoretical structure that is used for making truth-valued claims—such as predictions—about aggregate-level phenomena. Another option is to suggest that if claims about individuals are made, they must be qualified, resulting in ideas such as these: the model is only approximately true of real people; some of its assumptions are closer approximations, some others further away from the exact truth; some people are closer, some others further away from the model behaviour; and so on.

However, the idea of approximation alone is unhelpful. It must be put in a pragmatic context of purposes and interests. The question we must ask is this: *how close is close enough for serving this or that cognitive or non-cognitive interest or purpose?* How much divergence between the model and the target reality is *negligible* given the purposes sought to be served? To answer this question, we would need answers

to two other questions: What is the degree of approximation, the distance between the model and the target? What are the purposes that the model is supposed to serve?

The first of these questions—about the actual degree of approximation—is not easy to answer in the case of *homo economicus*; a sizable room for judgement remains. It should be easier to determine the purpose for which the model is supposed to be used. Such possible purposes include explaining aspects of individual behaviour; predicting aggregate level phenomena; and providing advice for policy and institutional design. Each such purpose must then be made more specific; for example, by specifying some required minimum degree of predictive precision, or by specifying some narrow aspect or feature of behaviour sought to be explained. Based on these two pieces of information—about the actual degree of approximation and the specific precise purpose—we could then decide whether the distance between what the model says (or what its assumptions appear to say) is the case and what actually is the case is negligible or non-negligible, that is, whether the model (or its assumptions) is close enough or too far from the facts of the matter (on the notion of negligibility, see Musgrave 1981; Mäki 2000, 2011).

The notion of negligibility can then be incorporated into that of *applicability*. For a model to be applicable to a domain of phenomena, it is not required that it is exactly accurate about the facts. It is enough if the extent to which it gets the facts wrong is negligible given this or that purpose. The notion of negligibility also helps to see the nature of those situations that prompt modellers to modify their models by relaxing idealizing assumptions and replacing them by more “realistic” ones. In such situations, the gap between a model and the target domain is considered too large to be negligible, hence the gap must be narrowed for it to become negligible for some purpose; thus (at least some of) the idealizing assumptions play the role of *early-step* assumptions that are to be relaxed and be replaced by less idealizing assumptions (again see Musgrave 1981; Mäki 2000, 2011, 2020).

Idealizing assumptions are often used for promoting the *tractability* of a model, enabling the use of a given set of formal tools that facilitate its theoretical and computational treatment (Hindriks 2006; Mäki 2011). The extent to which such assumptions distort the facts may or may not be negligible. In case it is negligible, there is no need for revision as long as a given purpose is being served. In case it is not negligible, there is need for relaxing the tractability-enhancing idealizations, but getting there may have to wait until other formal tools have been employed or developed. No doubt issues of tractability have played a role in the debates and developments around *homo economicus*.

### 3 Sources and Species of Pressure

What I call pressure is a loose metaphor that is intended to cover various constraints and demands—desiderata and criticisms—that often give rise to response. They may consist of theoretical goals and epistemic virtues, empirical evidence, challenges and

demands from outside, such as from other disciplines and commonsense points of view, and social reality.

### 3.1 Economics

Consider first *economics* itself. Why is it that one or another version of *homo economicus* is almost regularly assumed when economic theories and models are being constructed? Why is this? What purposes are being served? What pressures are being recognized and responded to, by yielding or by evading? Note that economics has no need for sticking to any conception of “human nature” (as is often suspected in critical commentaries). It does not hold such a conception, thus cannot be blamed for having got human nature wrong (more on this below). Hence there is no pressure deriving from a need—no demand—for a view of human nature. So what is the point? Why assume *homo economicus* if not attempting to capture human nature?

We need to understand the *disciplinary conventions* that drive and shape what economists do when constructing and accepting their models—they act under the pressure of such conventions, as it were. There are roughly two kinds of such pressure, call them substantive and methodological. *Substantive pressure* derives from disciplinary conventions that determine theoretical needs such as being able to derive downward sloping demand curves and to establish the possibility of, and some sort of attraction towards, market equilibrium. Merely for such derivational accomplishments, rationally choosing *homo economicus* is not necessary, as the same outcomes can be generated by assumptions about irrational or random choice, as established by Becker (1962). Yet some of the methodological conventions, such as those cited next, speak for the former.

*Methodological pressure* derives from disciplinary conventions that fix a set of methodological virtues and vices that shape and constrain what is being done and how, what sorts of models and assumptions to consider and accept, and what not. Here I mention just two such virtues. *Tractability* is a prerequisite that supports simplicity pragmatically conceived: simplicity of use. Tractability is a ubiquitous desideratum in any cognitive effort, but it takes on different forms in different settings, shaped by the concepts and formal tools available in a given research field. Simplicity—that contributes to tractability—is also a component in another methodological virtue, that of *unification*. Models and their assumptions are expected to explain much by little, that is, to unify a broad range of types of phenomena by deriving their descriptions from a small set of simple explanatory principles. Unification occurs in terms of simple *explanantia* that are stable across applications. It contrasts with complex *explanantia* that are ad hoc, varying from case to case (on unification in economics, see e.g. Mäki 2001; Mäki and Marchionni 2009).

No doubt *homo economicus* satisfies the desideratum of tractability, and it has usually been constructed and put in use with unification in mind. Tractability and unification can be seen as virtues of cognitive economy and efficiency. On the other hand, models of satisficing behaviour (in the vein of Herbert Simon), species of *homo*

*altruisticus*, or the varied models suggested in recent behavioural literature could appear inferior in not perfectly meeting these specific methodological pressures.

### 3.2 Common Sense

We then briefly discuss *the common sense* as a perspective from which to view *homo economicus*. From this perspective pressure keeps generated spontaneously. The source of this pressure consists of things such as ordinary everyday experience and reasoning, including introspection, commonsense concepts and beliefs. The first point to make is that, from the commonsense point of view, we tend to recognize *homo economicus* as being *about* human beings. It appears to *refer to* human agents and to attribute to them properties that are among the sorts of properties, roughly, that we humans have anyway—preferences and intentions, beliefs and expectations, etc. We recognize these items as familiar and talk about them as part of our ordinary everyday lives. In this sense, *homo economicus* does not go beyond the realm of *commonsensibles* (see Mäki 1998 for this notion).

That *homo economicus* and the common sense deal with the shared realm of commonsensibles facilitates the latter to impose critical pressure on the former. There indeed appears to be a rather stark conflict between the two. The conflict is not one of *referring to*, being *about* a shared domain, it is rather one of giving different *representations* of what they both are about, associated with respective beliefs. Commonsense representations are accompanied by *commonsense beliefs* about human behavior, and these appear to be at odds with the idealizations that make up *homo economicus*.<sup>1</sup> The relevant commonsense beliefs consist of what we ordinarily consider we humans are like, what we expect from ourselves and others, predictively and normatively. In an apparent contrast to *homo economicus*, the common sense recognizes divergent features such as plurality of clashing motives; radical shortage of information; limits in calculative powers, even reluctance to engage in calculation; weakness of the will; relatively strong other-regarding motives; limits to greed; importance of emotions, morality, community, reciprocity, fairness.

This clash is reflected in Philip Pettit's distinction between the common mind and the economic mind (Pettit 2001). Both are concerned with what we humans do, are inclined to do, and are advised to do. According to *the economic mind*, you do and should do *X* since it is in your personal advantage to do *X* (this captures the self-seeking dimension of *homo economicus*). According to *the common mind*, you do and should do *X* due to reasons such as membership in a group, loyalty, fairness, trust, kindness, reciprocity, honesty (as can be seen, this suggests two senses of 'common' in 'common mind'). As Pettit puts it, *homo economicus* "files in the face of common sense" that does not recognize the model as accurate of human beings

---

<sup>1</sup>Note that a more refined discussion would have to consider the asymmetry between commonsense beliefs and the assumptions of *homo economicus*: the latter are not beliefs, nor are they believed true!

in most contexts. I will return to Pettit's thoughts in the next section, to discuss the conditions of applicability of the model.

### 3.3 *Other Disciplines*

There is a long history of challenging *homo economicus* by appealing to research results produced by *scientific disciplines other than economics*, such as experimental psychology, cognitive neuroscience, anthropology, and sociology (see e.g. Coats 1976). Much of the pressure on *homo economicus* deriving from such other disciplines shares a great deal with the pressures coming from the common sense, often providing theoretical articulations and systematic empirical support for familiar commonsense perceptions. The most recent wave of behavioural economics is no exception to this.

Focusing here on the critical pressures from the behavioural challenge, it points out cognitive and motivational departures in human behaviour from conventional versions of *homo economicus* (see e.g. Thaler 2000). Based on experimental and other findings, it is argued that *we are not that smart*—that is, not very rational in the more conventional sense. We are not very informed nor terribly capacitated in information processing. We suffer from “biases” such as overoptimism, overconfidence, framing effects, loss aversion, fears, manias, and other sources of foolishness. As for what drives us, *we are not emotionally and morally cold, nor predominantly selfish*. We are capable of altruism, pro-social preferences, empathy, guilt, shame, regret, pride, joy, envy, hatred, love, and other familiar emotional and motivational states also recognized by the common mind.

A series of studies across fifteen culturally varied small-scale societies produced an interesting body of evidence (Henrich et al. 2004). In its “hunt for *homo economicus*” the interdisciplinary project combined experimental and ethnographic approaches, and went beyond the usual studies that use university students as their subjects. Applying game experiments like the ultimatum game, dictator game, and public goods game, the project focused on the selfishness component in *homo economicus*. Even though this large study left many issues open, it firmly concluded that *homo economicus* is falsified by the evidence from all fifteen small-scale societies; and that there is a great deal of cross-cultural variation in the extent to which subjects in these societies deviate from *homo economicus*—variation that cannot be explained in terms of individual differences, but requires invoking differences in the social and economic structure of these societies.

Linking this last conclusion with what we said above about the grounds—disciplinary conventions—on which economics tends to stick to one idealized model (or a family of closely knit models), we can see that the cross-cultural heterogeneity in behavioural dispositions poses a challenge to one standard defence of *homo economicus*. This defence is put forth in terms of the universality and unifying role of *homo economicus*: no other idea of human behaviour is able to unify an equally wide range of phenomena in equally simple terms. The results of the “hunt” seem



to suggest that no such unification can go through since there is too little unity in human behaviour across situations. Relatively small deviations from the model may be possible to accommodate by arguments from negligibility, but if the deviations are large and if they vary greatly from society to society, the task will be more difficult. This gives rise to issues of applicability, to be discussed in the next section.

Finally, there is the feminist objection, arguing that *homo economicus* turns out to be a model of economic *man* at most, thus biased and lacking in generality. The characteristics attributed to the economic agent are, at most, if at all, properties found in males. Ridiculing Adam Smith's famous passage, "it is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from regard to their own interest", Nancy Folbre (2001, 9) states and asks, "it is not usually the butcher, the brewer, or the baker who fixes dinner, but his wife or mother. Does she act out of self-interest, too?" The objection is that *homo economicus* fails to capture important parts of economically relevant action, such as household work, caring behaviour etc. This too will cast doubt on the generality and unifying capacities of *homo economicus*.

## 4 Issues of Applicability

An obvious first move in responding to pressure is to argue that the model of *homo economicus* is just like any other model in that it *isolates* (with the help of various omissions and idealizations) a fragment of the whole set of human dispositions and capacities, with an associated belief that this isolated fragment is the (most) relevant one (see Mäki 1992). This response alone would be incomplete, thus it must be amended with further steps. Among other things, these will specify criteria of relevance in terms of domains and questions.

There are several options for such further steps in responding to pressure. Consider model M1 applied to domain D1 – designate this as  $\langle M1, D1 \rangle$  – and consider pressure that requires response. One way to respond is to revise M1, resulting in  $\langle M2, D1 \rangle$ , which is hoped to eradicate or alleviate the pressure by improving the model by de-isolation or by re-isolation (where de-isolation is a matter of supplementing the isolated items with some additional items, while re-isolation occurs by replacing one isolation by another). Call this the *revise-the-model response*. Another way to respond is to adjust the intended domain of application, while keeping the model intact. The result might be  $\langle M1, D2 \rangle$ . The idea is that while M1 has problems with domain D1, it will be more smoothly applicable to domain D2. Call this the *adjust-the-domain response*. This response is often accompanied by some further ideas about problems posed or questions asked about phenomena in any given domain. Thus one may respond to a pressure by moving from question Q1 to question Q2 while at the same time either sticking to domain D1 or switching to domain D2. The hope is that the model is more capable of dealing with Q1 than with Q2. This is the *adjust-the-question response*.

While the revise-the-model response is the one pursued by behavioural and other associated approaches, the latter two responses—adjusting domains and/or questions—have been popular in past efforts to align conventional *homo economicus* with the world. These moves are expected to make it easier to justify *homo economicus* when appropriately applied; that is, to make it easier to defend the negligibility of its factual inaccuracies.

One example is the recent debate over the relevance of behavioural evidence for assessing *homo economicus*. A precondition of critical pressures from behavioural research to be even considered is an internalist interpretation of *homo economicus*. The model must be read as having contents dealing with internal mental capacities, dispositions, and processes. These contents could then be targets of the challenge. Another precondition is to set agent behaviour as the appropriate *explanandum*; or at least to establish it as a potential difference maker—and thereby a relevant *explanans*—for macro-level phenomena. On these conditions, theoretical change is an appropriate response to critical behavioural evidence.

One can escape the need for theoretical adjustment in response to behavioural evidence, however, by not accepting those preconditions. One may adopt an externalist interpretation of *homo economicus*, not allowing for any implications about internal mental matters; and one may insist that it is not the task of the model to explain agent behaviour but rather to explain and/or predict aggregate phenomena. Behavioural critique would have no or little bite, and *homo economicus* would be protected by setting it beyond the reach of behavioural evidence (see Gul and Pesendorfer 2008; cf. Hands 2013).

Most of the current debates about *homo economicus* have concerned human individuals, their preferences and the pursuit of maximum utility, but the earlier history of economics has also focused on using the model for depicting business firms and their rational pursuit of maximum profits. It is here that we find other illuminating examples. These include Fritz Machlup's (1967) account of the proper application of various theories of the firm. This account provides an applicability argument that combines domains and questions. The argument seeks to justify the model of the perfectly competitive profit-maximizing firm that had been challenged in the famous "marginalist controversy" two decades earlier in the 1940s. The challenge appealed to empirical evidence directly related to the decision-making behaviour by business firms. Critics argued that evidence shows that business firms do not maximize their expected returns, *homo economicus* style. Firms do not set the marginal cost equal to marginal revenue, required by profit maximization. Machlup responds by providing an argument that builds upon the presupposition that such direct comparisons between models and evidence alone are irrelevant and that they must be supplemented with considerations about domains and questions. He argues that it is not the task of this theory to help explain firm behaviour.

Among the characteristics of the *domain of phenomena* to which the model of perfectly competitive firm is applicable, Machlup cites a sufficiently large number of business firms that function under sufficiently competitive situations. And when using the model of perfectly competitive profit-maximizing firm, the proper *explanatory question* to ask is not about how individual firms behave. It is about qualitative

changes in market level variables, such as price and quantity, in response to changes in the circumstances. By contrast, other theories of the firm, such as managerial and organisational theories, can be used for asking and answering questions about firm behaviour, also when the numbers are small and the situations are not very competitive.

The conditions of applicability of marginalist theory suggested by Machlup thus combine domains and questions. As features of the domain of relevant phenomena, there must be large numbers of firms, and these must act under competitive circumstances; and regarding the explanatory questions, they must be the ones cited above, concerning qualitative market level changes as the *explananda*. If these conditions are met, then the divergence from the facts of the idealizations and omissions of the theory is negligible, and the model can be applied.

Another perspective from which to specify the conditions of applicability of *homo economicus* sets the focus on the *types of agent* of behaviour. One could argue that Machlup's account has carved out the proper domain of application of *homo economicus*, placing it in *business behaviour rather than consumer behaviour* (subject to further conditions pointed out above). The reason would be the fact that business behaviour is often subject to severe competitive exigencies to a far larger extent than is consumer behaviour. There is a causal mechanism for the elimination of non-profit-maximizing firms of the sort that does not control the fate of consumers, not at least with equal force (see e.g. Satz and Ferejohn 1995). In case one holds an image of *homo economicus* that associates it primarily with individual human beings, this would give reason for some reordering.

Another argument focuses on the *kinds of decision* that are made by economic agents rather than who these agents are. The issue is about the sorts of decision that *homo economicus* is more likely to be adept for making. Behavioural economists have concluded that evidence suggests that *homo economicus* is more suited for representing *simple and recurrent decisions* than complex long-term decisions. This would be a relatively easier domain of application, and again reason for some reordering. Here is a summary statement:

Ironically, rational choice models might therefore be most useful in thinking about the simplest kinds of decisions humans and other species make — involving perceptual tradeoffs, motor movements, foraging for food, and so forth — and prove least useful in thinking about abstract, complex, long-term tradeoffs which are *the traditional province of economic theory* (Camerer et al. 2005, 55).

Yet another way to single out the proper domain of application of *homo economicus* is in terms of *the economic domain* in contrast to non-economic domains. An early example of this is JS Mill's conception. Defining *homo economicus* in terms of wealth seeking, Mill believes that it dominates the economic domain. This is to say deviations from it are negligible here. However, such deviations are not negligible in the non-economic domains of family, friendship, community life. Hence in these domains *homo economicus* is not applicable: "With respect to those parts of human conduct of which wealth is not even the principal object, to these Political Economy does not pretend that its conclusions are applicable." (Mill 1844, 98)

As I read it, Philip Pettit (2001) proposes a refined Millian view on these issues. By and large, the distinctions between the economic and non-economic domains, and between the economic mind and the common mind coincide. Yet he suggests that there is a special role for the economic mind to play also in the non-economic domain. Focusing on the element of self-interest in *homo economicus* in the economic and non-economic domains, Pettit examines issues of existence and efficacy. What kind of *modal existence and efficacy* are enjoyed by the two minds, the common mind and the economic mind? The options include *actual and virtual* existence and efficacy. Pettit's suggestion is that *in the economic realm*, the economic mind *actually* rules. This is to say the self-seeking disposition continuously drives behaviour as a matter of actual fact. *In the non-economic realm*, the situation is less straightforward. The suggestion is that the common mind, most of the time, *actually* rules. This is to say we are mostly driven, as a matter of actual fact, by other-regarding motives and reciprocity, feelings of community and fairness, and the like. This set of drivers is actually active in mostly taking care of everybody's interests, so there is no need to mobilize the economic mind. Yet the self-seeking economic mind plays a role all the time *virtually*. It permanently lurks in the shadows, so to speak, ready to be mobilized when needed. It is disposed to get actualized at times when the agent feels her interests are not being sufficiently served just by relying on the operation of the common mind. This may occasionally happen in the non-economic domain, while it continually happens in the economic domain.

Pettit's account can be considered a refined version of an old position within economics, at least since Wicksteed and Marshall. As I said earlier, economics has no need to make assertions about some deep "human nature" such as claiming that ultimately and predominantly, people are egoists by their nature, regardless of context. Economics needs no such assumption. The traditional position is that it is sufficient to assume that *when engaged in market exchange* the agent will not promote the interests of the other party in the transaction, but rather seeks to maximize her own net benefits, utility or net revenue. After exiting the market, the agent may use that outcome as she wishes—donate it for philanthropy, please her family, drink herself to ruins etc. These activities are not—or do not have to be—supposed to be governed by rational self-interest. On this view, the market constitutes the domain of application of *homo economicus*. Or using a little stronger wording with a metaphysical bent, the market is the natural habitat of *homo economicus*.

This last idea gives rise to piquant issues about the boundaries of markets and about arenas of action more generally. These will play important roles as we next discuss the fourth and final source of pressure, that of changing social reality itself. In contrast to pressures deriving from the common sense and scientific disciplines other than economics, some of these pressures support *homo economicus*. There are trends in social reality that may make it more hospitable to *homo economicus*.

## 5 Social Pressures

The ways in which larger social structures and human behavioural propensities are related to one another is a perennial issue in the social sciences. As such a grand issue, it remains unsettled, but piecemeal progress can be made and has been made in clarifying its details. This is the case with *homo economicus* too. The picture of how its propensities are connected to larger institutional and cultural structures is subject to ongoing research and debate.

Here I briefly point out three sources of pressure in social reality, all of them supposedly supporting or reinforcing attributes of *homo economicus*. They are the *norm of self-interest* in some cultures; the involvement of *money*; and *marketization* of society. Unlike the reverse pressures from the common sense and other scientific disciplines, these are pressures that buttress and encourage the existence and exercise of the qualities of *homo economicus*. If these socially and institutionally based tendencies were to materialize uninhibited, the model would become increasingly accurate and its inaccuracies would become increasingly negligible.

Psychologists have investigated self-interest not only as a driving motive of behavior, but also as a *self-reinforcing social norm* (e.g. Miller 1999). The norm is found to be particularly strong in some cultures, for example in USAmerica. For it to function as a norm, people are *expected* to pursue their own interests, in activities ranging from those involved in workplace relations to voting in elections and beyond; and apparently non-selfish behaviour is often *interpreted* as ultimately selfish. In such cultures, the prevalent rhetoric of self-interest reinforces self-interested behaviour and the interpretation of behaviour as self-interested. People are moved by a fear to be exploited if they do not actively seek their self-interest. Institutions are designed and constructed assuming that self-interest is dominant. Such cultures would be more suitable domains of application of—at least this one attribute of—*homo economicus*.

As a special case of self-interest functioning as a social norm, recent research and debate on economics students prompts exciting questions. Does the study of economics reinforce self-interest as a norm? Economics students have been compared to students in other disciplines and found to behave in more self-seeking ways (see e.g. Frank et al. 1993). Is this due to them having been exposed to *homo economicus* as a major part of their education, inducing them to internalize self-interest not only as acceptable but as expected? Has a modelling assumption been turned into a social norm? This theme is too large to be explored further here. It just illustrates the general idea in this section.

Another source of supportive pressure on *homo economicus* in social reality is the *presence of money*. Again, psychologists have shown that when exposed to money in various ways—when handling money, when primed with money, when there is talk of money—people’s behaviour changes, resulting in reduced helping behaviour; reduced compassion; a business decision frame with unethical outcomes; endorsement of unregulated free market and social inequality (see e.g. Vohs et al. 2006). This suggests that situations that involve the presence of money in suitable ways would be

more apt for inclusion in the domain of application of *homo economicus* than those that do not involve money.

Money is of course not a separate institution with consequences for human behaviour. It is an integral part of the institutional structures of *the market*. As I said earlier, the market is the natural habitat of *homo economicus*. Calculative self-interest is often considered appropriate for the arena of the impersonal market, both morally acceptable and practically functional. Indeed, relatively free from moral blame, agents who are closer approximations of *homo economicus* will do better in the market arena. This means there is functional support available in social reality for these closer approximations.

Among the mega trends of contemporary culture, the *marketisation of society* plays a significant role in the persistence of *homo economicus*. Growing proportions of our social lives and institutional arrangements are governed by market principles (see e.g. Sandel 2012). This is to say that what is happening is relative growth of the market arena as well as relative growth of conceiving arenas of action as markets. This has consequences for how we behave, should behave, and expect others to behave. If indeed behaviour in a *homo economicus* style is appropriate for the market arena, then the trend towards extensions of the domain of the market will also push towards what we may call *homo economisation* of people. In general, this is an example of how institutions and individuals may co-evolve.<sup>2</sup>

Part of the trend of marketization is spontaneous, while another part is deliberately *designed*. This may be direct or indirect. Institutions and organizations may be directly designed on the basis of market principles; what was not a market arrangement may become so by design. On the other hand, policies may be designed and implemented on the presupposition that the relevant behavioural responses will be similar to those in the market, thereby indirectly promoting market principles. In both cases—designing institutions and designing policies—one assumes *homo economicus*, namely that people are, or will become, negligibly close to it in their behavioural responses.

This means that when designing institutions and policies based on the *homo economicus* presumption, one will, as it were, play the game of those who are closest to the model in their behaviour, thereby reinforcing their calculative self-seeking inclinations. The presupposition is that agents respond to incentives and thus their motivations are predominantly external. Those who satisfy these presumptions feel the game played is theirs and respond more or less predictably to the policy moves. On the other hand, those further away from *homo economicus* will feel less comfortable about the policies that do not rely on their internal motivations (such as those

---

<sup>2</sup>The connection between commercial market institutions and *homo economicus* has been recognized by social thinkers long ago. Max Weber is among them, writing this: "... the historical peculiarity of the capitalistic epoch, and thereby also the significance of marginal utility theory ... for the understanding of this epoch, rests on the circumstances that ... under today's conditions of existence the approximation of reality to the theoretical propositions of economics has been a *constantly increasing* one." (Weber 1973 [1908], 33) Thanks to Wade Hands for directing me to this passage.

deriving from duty, honour, decency, empathy, pride and shame). They will feel pressure to become closer approximations to *homo economicus* with weakening internal motivations.<sup>3</sup>

The previous paragraph invokes the quantification issue in terms of sets and subsets of individuals. There is indeed no reason to rely on a high degree of uniformity between people in regard to the relevant properties. Some have more of them in higher degrees, others far less. This suspicion is supported by empirical research. A study of 446 Tokyo residents, using dictator games and prisoner's dilemma games, provides an illuminating example (Yamagishi et al. 2014). The study concluded that the population can be roughly divided into three classes. One is the class of *homo economici*, those that are selfish and rational; they are intelligent, consistent, planful, socially successful, and they enjoy high life satisfaction. These closest approximations to the model comprise about 7% of the population. Another class is what the authors call *quasi homo economici*, those individuals who are selfish but impulsive, so not terribly rational; they are imprudent, reckless, socially unsuccessful, and have low life satisfaction. Their share was 8.7% of the population. The remaining and by far the largest group consists of ordinary people with ordinary pro-social attitudes and workable minimum degrees of rationality, as well as reliance on community spirits serving their interests sufficiently well.

If the conclusions of the Tokyo study are anywhere close to the representative facts, we can see that policies presupposing and promoting *homo economicus* style behaviour are tailored for the characteristic responses of just a fairly small minority of people without imposing much pressure upon them. Pressure upon others—ordinary community members and *quasi homo economici*—may be considerable, and failure to yield to the pressure may have harmful consequences for (the mental harmony of) those people and for society. The policies are designed in such a way that they force everybody to play the market game which is the game tailored for *homo economicus* types. While these types feel like being at home, others ultimately don't.<sup>4</sup>

## 6 Conclusion: Pressures and Tensions

*Homo economicus* lives its long life under multiple pressures—including those from the academic discipline of economics; from the common sense; from disciplines other than economics; and from various facets in social reality. These partly pull in different directions. The disciplinary conventions of standard economics, with partial support from aspects of social reality, are fans, while the common sense and many

---

<sup>3</sup>At this point psychiatrists may enter the conversation and start sharing observations about growing mental issues among various cohorts of the population.

<sup>4</sup>As Michiru Nagatsu pointed out to me, none of this rules out the possibility of policies that both presume and pressure agents to act like *homo economicus*; and that will not materially benefit the closest approximations to it (e.g. incentivizing them to contribute to public goods).



research fields other than standard economics are among the foes of the model of *homo economicus*.

Major tensions appear to prevail between the pressures on *homo economicus*. On the one hand, there are powerful forces in present society that press towards a *homo economization* of people, supporting their *homo economicus* qualities and pushing them to become ever closer approximations to the model. Advocates of the model in economics would welcome this as it would provide extra assurance that they are on the right track, encouraging them to go on basing their explanations and policy advice on the model—which in turn would reinforce the trends in social reality. On the other hand, the common sense and many research fields outside and even inside economics question the adequacy of the model other than for some restricted purposes. They imply that there are limits to *homo economization*, that is, limits beyond which it is no longer healthy to yield to the social pressures.

Nothing is perfectly unambiguous however. Consider the recent intrusions of some psychological insights into economics and the establishment of behavioural economics within standard economics. This has been enabled by not revising the grand scheme of constrained optimization, but rather by incorporating mental nuance into the scheme, thereby relaxing some of the idealizations of conventional *homo economicus*. Parts of *behavioural welfare economics* seem even more loyal to the more conventional versions of *homo economicus*, construing it as a normative ideal without those relaxations. In their actual behaviour, people make *mistakes*, that is, they deviate from the principles of *homo economicus* that, nevertheless, should be reconstructed as the ultimate arbiters in welfare considerations. This is to say behavioural welfare economics "... models human beings as faulty Econs [*homo economicici*]. Its implicit model of human decision-making is that of neoclassically rational inner agent, trapped inside and constrained by an outer psychological shell. Normative analysis is understood as an attempt to reconstruct and respect the preferences of the imagined inner Econ." (Infante et al. 2016, 22) This perspective is a natural ally to the policies and trends in social reality that support the *homo economization* of us. These are to be celebrated for helping us to discover and unleash the neat "rational inner agent" in us (the "Econ within") that is being suppressed by the messy psychological shell.

The pressures are not constant, and neither are the tensions and partnerships between them. The way these tensions are resolved and these partnerships are renegotiated will shape not only the future of the model of *homo economicus* but also the future of the sources of these pressures.

## References

- Becker, Gary. 1962. Irrational behavior and economic theory. *Journal of Political Economy* 70: 1–13.
- Camerer, Colin, George Loewenstein, and Drazen Prelec. 2005. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43: 9–64.



- Coats, A.W. 1976. Economics and psychology: The death and resurrection of a research programme. In *Method and appraisal in economics*, ed. Spiro Latsis, 43–64. Cambridge: Cambridge University Press.
- Folbre, Nancy. 2001. *The invisible heart: Economics and family values*. New York: New Press.
- Frank, Robert H., Thomas D. Gilovich, and Dennis T. Regan. 1993. Does studying economics inhibit cooperation? *Journal of Economic Perspectives* 7: 159–171.
- Gul, Faruk, and Wolfgang Pesendorfer. 2008. The case for mindless economics. In *The foundations of positive and normative economics*, ed. A. Caplin and A. Shotter, 3–39. Oxford: Oxford University Press.
- Hands, D.Wade. 2013. Foundations of contemporary revealed preference theory. *Erkenntnis* 78: 1081–1108.
- Henrich, Joseph, et al. (eds.). 2004. *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press.
- Hindriks, Frank. 2006. Tractability assumptions and the Musgrave-Mäki typology. *Journal of Economic Methodology* 13: 401–423.
- Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden. 2016. Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* 23: 1–25.
- Machlup, Fritz. 1967. Theories of the firm: Marginalist, behavioral, managerial. *American Economic Review* 57: 1–33.
- Mäki, Uskali. 1992. On the method of isolation in economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 26: 319–354.
- Mäki, Uskali. 2000. Kinds of assumptions and their truth: Shaking an untwisted F-twist. *Kyklos* 53 (3): 303–322.
- Mäki, Uskali. 2001. Explanatory unification: Double and doubtful. *Philosophy of the Social Sciences* 31: 488–506.
- Mäki, Uskali. 2012. The truth of false idealizations in modeling. In *Models, simulations, and representations*, ed. P., Humphreys, and C. Imbert, 216–233. London: Routledge.
- Mäki, Uskali. 2020. Puzzled by idealizations and understanding their functions. *Philosophy of the Social Sciences* 50(3): 215–237.
- Mäki, Uskali, and Caterina Marchionni. 2009. On the structure of explanatory unification: The case of geographical economics. *Studies in History and Philosophy of Science* 40: 185–195.
- Mill, John Stuart. 1844. On the definition of political economy; and on the method of investigation proper to it. In *Essays on some unsettled questions of political economy*. London: Longmans.
- Miller, Dale T. 1999. The norm of self-interest. *American Psychologist* 54: 1053–1060.
- Morgan, Mary. 2006. Economic man as model man: Ideal types, idealization and caricatures. *Journal of the History of Economic Thought* 28: 1–27.
- Musgrave, Alan. 1981. “Unreal assumptions” in economic theory: The F-twist untwisted. *Kyklos* 34: 377–87.
- Pettit, Philip. 2001. The virtual reality of *homo economicus*. In *The economic world view. Studies in the ontology of economics*, ed. Uskali Mäki, 75–97. Cambridge: Cambridge University Press.
- Sandel, Michael J. 2012. *What money can't buy: The moral limits of markets*.
- Satz, Debra, and John Ferejohn. 1994. Rational choice and social theory. *Journal of Philosophy* 91: 71–87.
- Thaler, Richard H. 2000. From *homo economicus* to *homo sapiens*. *Journal of Economic Perspectives* 14 (1): 133–141.
- Vohs, Kathleen D., Nicole L. Mead, and Miranda R. Goode. 2006. The psychological consequences of money. *Science* 314: 1154–1156.
- Weber, Max. 1975 [1908]. Marginal utility theory and ‘The Fundamental Law of Psychophysics’. *Social Science Quarterly* 56: 21–36.
- Yamagishi, Toshio, Yang Li, Haruto Takagishi, Yoshie Matsumoto, and Toko Kiyonari. 2014. In search of *homo economicus*. *Psychological Science* 25: 1699–1711.