ERASMUS JOURNAL FOR PHILOSOPHY AND ECONOMICS
VOLUME 8, ISSUE 2, AUTUMN 2015

The Erasmus Journal for Philosophy and Economics (EJPE) is a peer-reviewed bi-annual academic journal supported by the Erasmus Institute for Philosophy and Economics, Faculty of Philosophy, Erasmus University Rotterdam. EJPE publishes research on methodology of economics, history of economic thought, ethics and economics, and the conceptual analysis of inter-disciplinary work relating economics to other fields. EJPE is an open-access journal. For additional information, see our website: <http://ejpe.org>. All submissions should be sent via e-mail to: <editors@ejpe.org>

EDITORS
Willem J. A. van der Deijl
Osman Çağlar Dede
Joost W. Hengstmengel
Luis Mireles-Flores
Philippe Verreault-Julien
Thomas Wells

ADVISORY BOARD
Erik Angner, Roger Backhouse, Marcel Boumans, Richard Bradley,
Nancy D. Cartwright, David Colander, Job Daemen, John B. Davis,
Sheila C. Dow, Till Grüne-Yanoff, D. Wade Hands, Conrad Heilman,
Frank Hindriks, Clemens Hirsch, Geoffrey Hodgson, Elias Khalil, Arjo Klamer,
Alessandro Lanteri, Aki Lehtinen, Uskali Mäki, Caterina Marchionni,
Deirdre N. McCloskey, Mozaffar Qizilbash, Julian Reiss, Ingrid Robeyns,
Malcolm Rutherford, Margaret Schabas, Eric Schliesser,
Esther-Mirjam Sent, Robert Sugden, Jack Vromen.
## TABLE OF CONTENTS

### ARTICLES

- Collective intentionality and the state theory of money  
  **GEORGIOS PAPADOPOULOS**  
  [pp. 1-20]

- Methodology in *Capital in the twenty-first century*:  
  a “new-historical” approach to political economy  
  **LUKE ANTHONY PETACH**  
  [pp. 21-33]

- A welfarist critique of social choice theory:  
  interpersonal comparisons in the theory of voting  
  **AKI LEHTINEN**  
  [pp. 34-83]

### BOOK REVIEWS

- William Davies’s *The happiness industry*:  
  *how the government and big business sold us well-being*  
  **JEFFREY R. DI LEO**  
  [pp. 84-93]

- Dani Rodrik’s *Economics rules: the rights and*  
  *wrongs of the dismal science*  
  **N. EMRAH AYDINONAT**  
  [pp. 94-101]

- Marcel Boumans’s *Science outside the laboratory*:  
  *measurement in field science and economics*  
  **ALESSANDRA BASSO**  
  [pp. 102-110]

- Lawrence A. Boland’s *Model building in economics*:  
  *its purposes and limitations*  
  **JAACKO KUORIKOSKI**  
  [pp. 111-116]

### PHD THESIS SUMMARIES

- Reconciling normative and behavioural economics  
  **GUILHEM LECOUTEUX**  
  [pp. 117-120]
The world as a garden: a philosophical analysis of natural capital in economics

C. Tyler Des Roches

[pp. 121-123]
Collective intentionality and the state theory of money

GEORGIOS PAPADOPOULOS
Aarhus University
University of Applied Arts Vienna

Abstract: The circulation of non-convertible currency and the source of its value raise important ontological questions that touch upon the conditions of its acceptance. The aim of this paper is to address such questions by illustrating how collective intentionality and constitutive declarations can be employed in order to develop an adequate ontological framework for explaining the emergence and the persistence of the current monetary standard. This analysis of money differs from that of mainstream commodity theory in that it argues against individualism, which traditionally underwrites both economic and philosophical analyses of money. The resulting ontology is based on an account of collective intentionality developed upon the “sharedness” of individual intentional states; this account supports the state theory of money, combining it with an ontological analysis of the state and its authority.

Keywords: collective intentionality, commodity theory of money, individualism, institutions, state theory of money

JEL Classification: B52, E40

THE IMPORTANCE OF THE ONTOLOGY OF MONEY

The ontology of money has been a fertile field for the application and the appraisal of philosophical theories of social existence; many social philosophers including Karl Marx, Georg Simmel, and Max Weber have developed elaborate theories about the emergence of money and its acceptance in economic transactions. More recently, the ontology of money has been an important question in the debate on the usefulness (or not) of “collective intentionality” (Gilbert 1989; Searle 1995), and further, has been used an illustration for the usefulness of the notion in the analysis of social existence (Hédoin 2013; Mäki 2005; Searle 1995, 2005).

Author’s Note: The author would like to thank the editor of this journal, Philippe Verreault-Julien, as well as two anonymous referees for their helpful comments.
The aim of the paper is twofold:

1. To contribute to the debate about collective intentionality by illustrating the explanatory merit of the notion for the social existence of money. This identifies money in the context of the current system of fiat money where its value is not intrinsic (or “hylic”) and the state serves an important function as the ultimate guarantor of monetary value (Knapp 1924, 4). The proposed ontological account of money will update the state theory following the recent developments in social ontology.

2. To develop an account of collective intentionality built upon the notion of “sharedness”\(^1\) in order to propose an ontological framework for the state theory of money, explaining the preconditions for the social constitution and the mechanism of the collective acceptance of fiat money.\(^2\)

The state theory argues that money is essentially created by the state, where currency is enacted by declaration independently of the inherent value of the objects used to represent currency. The state's reliance upon money and its fiduciary nature is consistent with its current monetary regime; non-convertible currency that is issued by the central bank—a state institution that may or may not be independent of the government—is intrinsically worthless. The central bank infuses the system with trust because it guarantees both the acceptability of money and the security of deposits by assuming the function of lender of last resort. According to the state theory, money falls back upon the rule of law, the system of taxation, and the ability of the state to enforce laws and taxation.

The appeal to state authority provides only a provisional explanation for the existence of money, i.e., for its instantiation in objects and practices. An analysis of the ontology of intrinsically worthless and non-convertible money enacted by state authority should also provide an account of how this authority came to exist, and how it relates to the

---

\(^1\) The description of collective intentionality as shared intentionality comes from Schmid (2003). It is different from Searle’s account, which describes collective intentionality as the first-person plural-form of intentionality (Searle 2010, 43). Schmid reflects: “It is by now a well-established fact that intentionality is not exclusively a matter of the personal beliefs, desires and expectations of individuals. What makes our intentionality and our actions social is not just that from time to time, we make each other the object of our individual intentions or expectations. Rather, intentionality is in itself something human beings can share” (2003, 203; italics in the original).

\(^2\) Mäki (2005) and Tieffenbach (2010) provide examples of an ontological analysis of money that also remains attached to methodological individualism.
action and the attitudes of the individuals it regulates. Social ontology can provide the basis for the study of fiat money delineating a form of collective acceptance that is both able to carry the weight of fiat money and is consistent with a wider framework of analysis of social institutions, including the state. An account of collective intentionality that is based on the sharedness of collective intentional states—of which cannot be reduced to an aggregate of individual intentional states—can support both the existence of fiat money as well as the state that sanctions it. Here, collective intentionality is supplemented by the concept of “constitutive declaration” which, I argue, can resolve questions that concern the ontological foundation of the state theory of money—viz. where its institutional status as legal tender (not its value as a commodity) allows for its function.

In what follows, I elaborate the concepts of collective intentionality and constitutive declarations, highlighting their importance for the existence of social facts and for fiat money in particular. The analysis focuses on the relational character of collective intentionality, and on the explanatory advantage it provides over weaker notions of collective acceptance that are associated with methodological individualism (Meijers 2003; Schmid 2003). Furthermore, I illustrate the deficiencies of the individualist commodity theory, especially in relation to the current monetary standard where the acceptance of the means of exchange is not backed by a commodity guarantee. The problem of “group mind” is also addressed because it offers one of the main arguments against ontological collectivism (Searle 1990, 25). Finally, the proposed ontology of fiat money can account for the contribution of the state in the constitution and support of money, effectively filling an important explanatory gap within the state theory thereby countering the claim

---

3 Institutions can be defined "as systems of established and prevalent social rules that structure social interactions" (Hodgson 2006, 2).
4 Constitutive declarations are a special kind of speech act introduced by John Searle. Constitutive declarations constitute new social facts by representing speech acts as existing (Searle 2010, 93). The term is going to be further developed in the next section.
5 Schmid states that, “[w]here there is intentionality, it is said, there has to be somebody who ‘has’ it—the good old subject. Now if it is claimed that there is such a thing as collective intentionality, and that collective intentionality has to be distinguished from individual intentionality, the conclusion seems to force itself on us that it has to be not the single individuals, but the collectives themselves that ‘have it’. And for collectives to have intentions, some sort of a ‘collective mind’, some ‘group mind’ seems to be required, something hovering over and above the mind of the individuals involved” (Schmid 2003, 214).
that the state and its operation is just an all-purpose and ad hoc explanatory variable.

**BASIC CONCEPTS FOR THE ANALYSIS OF SOCIAL EXISTENCE; SOCIALITY, COLLECTIVE INTENTIONALITY, AND CONSTITUTIVE DECLARATIONS**

The starting point of an ontological analysis of money is the distinction of types of existence that characterize the social and the natural world, introduced by John Searle in *The construction of social reality* (Searle 1995, 5-13). According to Searle there is a fundamental distinction between natural facts and social facts. Natural facts “do not need us in any way”—their existence is independent of our representations about them. Social facts, on the other hand, are dependent on human consciousness for their representation. What defines sociality and social facts is a shared meaning supported by a shared language; the representations that we share about human interaction brings social phenomena into existence.⁶

Shared meaning is crystallized in collective intentionality: it is this collective intentionality that founds the proposed ontology of money, and thus explains the emergence and persistence of sociality and social facts. Intentionality is a broad philosophical concept that denotes more than just intention. It refers to the relation of the mind to the world, a relation towards external objects, states of affairs, and ideas. Candidates for intentional states can be any kind of mental representation of the world; representations are always about something or in reference to something (Searle 2010, 24). The relation of these mental representations to the world forms the basis of human consciousness, of human action, and allows for the constitution of social facts—this includes money. Collective intentional states employ the first-person plural form, meaning that collective intentional states express a “we-mode” rather than the “I-mode” that characterizes individual intentionality (Searle 2010, 47). The first-person plural form places the individual’s intention in relation to the collective of individuals to which collective intentionality applies (Davis 2003, 131). Collective intentionality here is understood as a shared we-intentionality and not as just an aggregate of individual intentional states; it is a particular type of intentionality that expresses an individual conviction and

---

⁶ Searle writes, “[t]he key to understanding intentionality, at least for these simple cases, is representation in a very specific sense. The intentional state represents its condition of satisfaction” (2010, 29).
participation in an intentional state that is shared by a collective of individuals. The members of we-intentionality share we-intentions as a collective, and the relations of the individual parties of collective intentionality—as parties that share the same we-intention—are integral to the content of we-intention. The relational character of the intentional states suggests a stronger, i.e., more collectivist, account of collective intentional states than Searle’s, who subscribes to a subjective notion of collective intentionality (Searle 1995, 26). The simple example of two people going on a walk together can illuminate the relational character of attitudes that constitute collective intentionality. The shared we-intention of each of the two individuals that go on a walk makes sense only if they both share the same we-intention to go for a walk. The shared intention creates a relation between the two individuals, and it is because of this relation that they go for a walk together. They act and perceive of their action as part of a common enterprise that they (both) acknowledge (Gilbert 1990, 7).

Collective intentionality is defined here in terms of shared meaning. Shared meaning founded on collectively recognized representations provides the foundation for social facts and for institutions, including money. We constitute the social environment on the basis of the shared representations and use these representations as the basis of our social interaction. The act of representation, when it is collectively acknowledged, generates new meanings and these meanings create new social facts that inform and condition human interaction. To return to the question of the ontology of money, money exists because we share representations that establish facts about meaning and the function of money—this is what gives it social significance in our everyday interactions. Social constitution can then be defined on the basis of a simple principle: social facts are established through our shared representations of them as existing, while these representations ascribe and reinforce the social significance of these facts.

The performative function of language, its ability to transform the world by ascribing meaning to facts, is presupposed by social constitution. John Austin developed a coherent account of how we can “do things with words” (Austin 1962). Searle expanded on Austin’s theory of speech acts (Searle 1969) and introduced the concept of constitutive declarations. Constitutive declarations are a particular type
of speech act that constitutes new social functions and social facts. Through these constitutive declarations we constitute social facts by declaring them to be the case. Constitutive declarations are public and official statements; they communicate the social significance of constituted facts and signal the authority that enacts these facts. Common examples of constitutive declarations include pronouncements of marriage and divorce (e.g., “I declare you husband and wife”), declarations of war, commencement and conclusion of sessions in official bodies such as court or parliament, and the like.

Constitutive declarations can be formalized as: “We (or I) make it the case by declaration that the Y status-function exists” (Searle 2010, 93). The formalization suggests that constitutive declarations need to be public and that their intended outcome is the social constitution of status-function. The public character of the constitutive declaration is important because the constituted status-functions need to be known (or at least knowable) to all the members of the collective, where the new status-function is enacted. Public acknowledgement does not entail that the constitutive declaration has to be performed verbally, or that it needs to be performed at all. Constitutive declarations are often formal rules that are made public by the appropriate authorities and are knowable, at least in principle, by all the affected parties. Status-functions are defined by a collectively acknowledged social representation, which in turn allows the fulfillment of a function. They are constituted by collective intentionality and they function by carrying deontic powers. Regarding the status of money, something underwrites the means of exchange, thereby enabling the process and fulfillment of economic transactions. New status-functions need to be acknowledged and endorsed by the collective, where the act of constitution is performed, in order to be effective. Only if the new representation that constitutes the status-function is shared by the collective will the declaration become successful. The success of constitutive declaration depends upon both public acknowledgment and the legitimacy of the declaration.

The legitimacy of constitutive declaration is supported by political authority, which is built upon an institutional structure that enables the

---

7 Meijers states that, “[t]he remarkable consistency of Searle’s project is also evident from the fact that social reality has been on his mind from Speech Acts to The Construction of Social Reality (1995). One of the revolutionary aspects of his theory of speech acts was the idea that speaking is acting in accordance with social rules. These rules not only regulate but also define linguistic utterances” (2003, 170).
performance and enforcement of such constitutive declarations. The constitution of status-functions presupposes the existence of an institutional structure that regulates the process and defines the conditions for the enactment of constitutive declarations (Searle 2005, 9-10). A special authority is not only necessary to legitimize the constitutive declaration, it is also necessary for the enforcement of the consequences of the constitution of a new status-function. The special status that enables social constitution also contributes to the investment of the new status-functions with collective intentionality. For example, only the government (or only the president) can declare war, while only a public official or a priest can legitimize marriage or divorce. These two conditions, public recognition and proper authority, are necessary but not sufficient for the establishment of a new status-function; it is the collective intentionality of the public that enables the constitution of status-functions. Constitutive declarations articulate and communicate the content (denoted by shared linguistic-mental representations) of the new status-function that they establish; but, it is only if the public collectively intend the new status-function, will the act of constitution be successful.

Collective intentionality and the status-functions that collective intentionality supports are the outcome of the negotiation of agents with different viewpoints and interests that try to promote a particular representation of social reality. Authority can then be conceptualized as the ability to tap into the institutional and discursive power structures in order to impose new representations of reality, universalizing and effectively constituting reality. Collective intentionality is both the precondition and the outcome of such a universalization. The power to enforce constitutive declarations and secure the collective intentionality of the public translates as the power to enforce social facts. The stake of social antagonism is to constitute partisan viewpoints as the universal interpretations of social reality. Authority is necessary for the constitution of money; it inspires and reinforces the collective intentionality towards money, which underlies its emergence and persistence, and it aligns the expectations of all individual users toward a general acceptance of the dominant standard of abstract value.

**MONEY AND COLLECTIVE INTENTIONALITY**

Searle has made extensive use of money to illustrate his account of social ontology (Searle 1995, 2005, 2010). Even in the introduction of
The construction of social reality (Searle 1995) money provides the starting point for the explication of the Searlean ontological framework. Money remains the focal point of debates about collective intentionality in the work of both proponents and critics of Searle’s account of the ontology of money. The latter have looked into mainstream economics in an attempt to find counter-arguments against the relevance of collective intentionality for social constitution against claims of the irreducibility of collective intentionality, and against the two-tiered ontology that distinguishes the social from the natural (Smit, et al. 2011; Tieffenbach 2010). This paper, then, can be seen as a criticism of Searle—the analysis of collective intentionality vis-à-vis the institution of money is intended to identify a stronger account of collective intentionality and constitutive declarations that is built upon the idea of “sharedness” (Schmid 2003). The aim is to develop an ontological account of money that makes the state theory consistent with recent developments in social ontology (Papadopoulos 2009).

Money should not be conflated with the objects that represent money, be it commodities or intrinsically valueless currency (coins or notes), since the institution of money depends on the shared representations (the speech-enacted meanings) that define its institutional status. Currency is used as a means of payment only because of its status-function as money; its status-function informs the very attitudes and behaviors about it. It is thus by virtue of its status-function that money fulfills its institutional role. The invisible hand explanation of the emergence of money (Menger 1892) can be reformulated by using the format of constitutive declarations; a certain commodity becomes institutionalized as money when it assumes the status-function of a medium of exchange. A similar formulation can be constructed for the state theory of money (Knapp 1924). Legal tender issued by the political authority becomes money when they assume the status-function of a standard of value. Yet, at a specific point in time, those commodities—such as pre-weighted pieces of metal or bills of exchange—stop being used as and become standards of value and media of payment because they assume status-function. The political authority enacts and communicates the status-function of currency by inscribing an insignia to it, by supporting it with its power, and by enacting the

---

8 Searle writes: “For example how can it be a completely objective fact that the bits of paper in my pocket are money, if something is money only because we believe it is money?” (1995, 2-3).
necessary legal rules. All of these steps serve to enforce the collective intentionality of the users.

As explained above, money can be successfully constituted only when it is enacted and communicated by agents who have the right to do so. The right to issue money is reserved for the sovereign political authority, who guards the monopoly over the monetary system through legislation and the use of force. Political authority represents the community it governs, and the monopoly to issue money is exercised in the name of this community. The officially issued currency carries the insignia of this authority, which Searle refers to as “status indicators” (Searle 1995, 119)—these symbols represent the status-function of the currency as the official legal tender. The symbolism of the insignia is intended to communicate the status-function of money, communicating that money will be accepted as payment, including the settlement of taxes. Authoritative support, which is expressed both symbolically and legally, injects money with the collective intentionality that it enjoys both as the issuer of money and more generally as the sovereign political agent within the community.

THE INSUFFICIENCY OF INDIVIDUALISM AS THE ONTOLOGICAL FOUNDATION OF THE CURRENT MONETARY STANDARD

One of the main problems of the commodity theory of money is its commitment to individualism (both ontological and methodological); by consequence it is unable to explain the emergence and persistence of money in the absence of a commodity guarantee. The individualistic methodology was established as the mainstream approach during the Methodenstreit (which translates as ‘conflict about methods’). During this time Menger published his famous article on money (Menger 1892), which subsequently influenced debates between the proponents of the commodity theory of money and the state theory of money. Commodity theorists argued that the use of money and the (indirect) exchanges of goods are the outcome of the individual maximizing behavior. The rational reconstruction of the emergence of money offered by Carl Menger (1892), along with the expansion of his model in neoclassical economics, has provided a consistent account of one possible, albeit historically and anthropologically unsupported, mechanism for the constitution of commodity money as the universally accepted means of exchange. The standard invisible hand explanation may be sufficient for

---

the study of commodity money, but it is not readily applicable in the current monetary standard, where money is not backed by a commodity guarantee. The Mengerian account was not intended to, and cannot, explain why individuals will accept worthless, token-money in exchange for valuable commodities, or further, how such a monetary standard could emerge. An intrinsically worthless means of exchange is vulnerable to the “free rider” problem because, as long as it remains intrinsically valueless, individuals will be better off if others exchange their goods for intrinsically “worthless” money with them, while they exchange their goods only for other goods. The positive feedback (viz. the reduction of transaction costs by a commonly accepted means of exchange that instigates the emergence commodity money according to the theory) is canceled out when there is no commodity guarantee, so a means of exchange cannot arise spontaneously. Further, even if there were a case in which a universal medium of exchange was not backed by a commodity (say through the suspension of convertibility in a pre-existing commodity standard) its persistence would remain a problem without the existence of an external mechanism to ensure compliance.

One solution to this problem is presented by Nobuhiro Kiyotaki and Randall Wright (1989, 1991, 1993), who argue that for a fiat money equilibrium to persist it is sufficient that each agent believes that all other agents will continue to accept fiat money for the commodities they want to exchange. The problem is that the postulated collective acceptance is not resilient if it is understood as an aggregate of individual beliefs and thus cannot be reconciled with methodological individualism. Indeed, Kiyotaki and Wright presuppose an individual belief about a collective belief (“I believe that everybody believes that money is and will remain acceptable”). Nevertheless, the individual belief of every agent is predicated on and conditioned by the collective belief, the status of which is contested in the framework of

---

10 There are two ways to facilitate fiat exchange in such a setting. Either “(1) by imposing a boundary condition, or (2) avoiding the boundary condition by pushing it away to infinity. Both are devices to circumvent the unraveling of the monetary equilibrium through backward induction” (Kovenock and De Vries 2002, 147). The boundary condition can be a policing authority or the assumption that individuals will continue to accept money, come what may.

11 Kiyotaki and Wright note that, “[t]o this end, we now suppose that everyone believes that others will accept fiat money and ask if this could be an equilibrium” (1989, 493; emphasis in the original). This is a familiar strategy in overcoming the problem of explaining the value of fiat money by assuming it. Sidrauski, in one of the first attempts to incorporate money in general equilibrium modeling, assumed that real cash balances yield positive utility (Sidrauski 1967, 535).
methodological individualism, especially since the beliefs of every individual are derived from, rather than constitutive of, this collective belief. Hence, the problem of the emergence of fiat money remains unsolved. Kiyotaki and Wright, and methodological individualists in general, are unable to answer the question of how such collective beliefs about the general acceptability of and faith in fiat money can arise and persist (remember that the function of fiat money as a means of exchange is undermined by free riding). Fiat money cannot emerge on the individual level, since the individual belief of each and every user of money presupposes a collective agreement about the acceptance of money by all other users. To put it more clearly, the establishment of the collective acceptance of money as merely an aggregation of individual beliefs is not possible because these individual beliefs presuppose the collective recognition needed to constitute it. The problem can only be circumvented with the postulation of a collective recognition of money shared by all individual agents in the market. The inability of methodological individualism to account for the emergence and persistence of fiat money, one of the most important economic institutions, indicates that the commitment to ontological individualism should be reconsidered, at least as far as the underlying philosophy of economic inquiry is concerned. In the next section I will defend a collectivist ontology based on collective intentionality and constitutive declarations. This counters any attempts to reduce the proposed ontological account to some kind of individualism, thereby strengthening the proposed framework for the explanation of the emergence and persistence of state sanctioned money.

THE IRREDUCIBILITY OF COLLECTIVE INTENTIONALITY

The foundation of the proposed ontology of fiat money, and more generally of social institutions, is a kind of collective intentionality based on “sharedness”—this is the strong collective intentionality that I have pursued thus far. As I argued above, such an account of collective acceptance is necessary for the persistence of a fiat money equilibrium.

---

12 Searle describes it as follows: “There is a deep reason why collective intentionality cannot be reduced to individual intentionality. The problem with believing that you believe that I believe, etc., and you believing that I believe that you believe, etc., is that it does not add up to a sense of collectivity. No set of ‘I Consciousnesses’, even supplemented with beliefs, adds up to a ‘We Consciousness’. The crucial element in collective intentionality is a sense of doing (wanting, believing, etc.) something together, and the individual intentionality that each person has is derived from the collective intentionality they share” (Searle 1995, 24-25, italics in the original).
In this section I am going to argue against attempts to reduce collective intentionality into individual intentionality, or to individual intentionality in conjunction with other individual attitudes (like common knowledge). By establishing that collective intentionality cannot be reduced to individual attitudes, the analysis of money emerges as substantially different from the account offered by the commodity theory; furthermore, it establishes how a genuinely collectivist attitude, like the proposed version of collective intentionality based on sharedness (argued for above; see Schmid 2003) is necessary for the emergence and persistence of the institution of money.

Collective intentionality requires at least two parties in order for we-intentions to be genuine. Sharing the same we-intention puts the parties in a relation, a relation that is itself part of the content of we-intentions. The fact that the parties of collective intention are in relation is indicated by the use of the first-person plural form of the shared intention; the “we-mode” of collective intentionality suggests that I see myself as part of a collective that intends in concert, and if the we-intention is correct, I am actually part of this collective of individuals intending together. The web of relations where the individual is placed by a collectively shared intention is internal to the collective intention because the propositional content of the we-intention describes the intention of a collective, and thus the content of the we-intention makes sense only in the context of the internal relations of this collective. To wit, collective intentionality and the relations of the individual parties are mutually constituted; the parties of the we-intention are related in virtue of the we-intention they share, and the propositional content of the we-intention makes sense only on the basis of these relations. If this claim is true, a reduction of collective intentionality to individual intention(s) is not possible because these essential relations will remain unaccounted for. The underlying conviction behind this claim is that it is the relations defined by and defining of collective intentionality that carry the act of social constitution.\(^\text{13}\)

The beliefs that underlie the existence and circulation of money can be reconstructed as an expression of the collective intentionality of the

\(^{13}\) Hodgson states, that “[t]he social world, by virtue of the fact that it is social, must involve such interactive relations. The term ‘social’ here is used in a broad sense, to encompass phenomena that are examined in economics, as well as other social sciences. In the social context all relations between individuals are causal and interactive, at least in the sense that in maintaining these relations with others, individuals are affected by their (partial) awareness of them and different actions may be enabled” (Hodgson 2007, 212).
users, e.g., “I believe that we recognize X as the standard abstract of value”, or “I believe that we accept X as a means of payment”. Such we-intentions cannot be reduced to an aggregate of I-intentions because they depend on a shared commitment to the acceptance of money, as well as to a wider framework of economic relations that define the market system of monetary exchange. These relations are part of the collective we-intention that supports money, not only because they constitute the market but also because they define the very notions of “standard of value” and “means of payment”. Money presupposes division of labor, markets, individual producers and consumers with specific attitudes and expectations about money and about the behavior of others towards money. The acceptance of the institution of money situates the individual in a web of economic relations with other producers, consumers and authorities that issue the money. These relations are what give money its social significance and identity.14

The content of the we-intention that each individual holds as a party involved in collective intentionality is derived by the content of the shared we-intention of the collective. Sharing collective intentions does not depend on mere individual awareness of the intentional states. Awareness of one’s sharing of a collective representation is not a sufficient condition for collective intentionality because one can mistakenly think that one is part of a collective intentional state that nobody else shares. If somebody we-intends to go for a walk with somebody else, the we-intention is justified if both partners intend to go for a walk; the propositional content of each individual's we-intention is dependent on and derives its validity from the propositional content of the shared we-intention. If we think correctly of ourselves as members of a collective sharing a collective intention, it is because we actually are part of this collective that shares this collective intention, and not the other way around (Schmid 2003, 212). Subjective individualism,15 the

14 Ingham clarifies that the “[m]onetary systems are the result of the long term historical development of a complex structure of social relations and practices which cannot be grasped by of neoclassicism’s methodology. In this respect, Smithin has observed that “the micro-foundations of standard monetary theory have been left extremely weak” (Smithin 1994, 14). In fact, we need to go further: money cannot have “micro-foundations” if these are sought exclusively in the formal deductive model of the individual agent’s rational choice of holding a “veil” or “lubricant” as simple medium in a “real” exchange economy” (1996, 516).

15 I shall use the term subjective individualism, following Hans Bernhard Schmid, in contrast to reductive or formal individualism, where all collective or holistic facts can and should be reduced to aggregates of individual attitudes or behavior. Schmid states: “Subjective individualism does not limit intentionality to the singular form, but
ontology that supports Searle’s weaker conception of collective intentionality, may lead to the opposite paradoxical conclusion that collective intentionality constitutes the collective that is expressed by the ‘we’ for every individual that holds a we-intention even in the case when the individual in question is the only one that holds the we-intention. The content and the validity of the individual we-intention is dependent on and derived from the shared we-intention. Individual beliefs about money presuppose a collective belief about money and the individual belief of every agent is predicated on and conditioned by the collective belief about the acceptability of money. Rephrasing the claim once more for the benefit of clarity, the constitution of the collective acceptance of money as the aggregation of individual beliefs of acceptance is not possible because individual beliefs presuppose the collective acceptance they are supposed to constitute.

Money is enacted through a constitutive declaration made by the sovereign political authority, and its constitution aligns the expectation of individual agents; this inspires a shared collective intentionality that underlies the individual we-attitudes towards money. Reductive or formal individualism, i.e., the attempts to reduce collective intentional states to aggregates of individual attitudes, is insufficient because the relations between the individual bearers of collective intentionality cannot be included into individual intentionality. Similarly, subjective individualism, which suggests that collective intentionality could exist in isolated individual minds, is also untenable. Furthermore, the internalist account of collective intentionality is untenable because collective intentional states (we-intentions) rely for their validity on a collective intention that is shared by the other parties.

Collective intentionality, like all other types of intentionality, is part of the spectrum of conscious states; so if collective intentionality is not reducible to individual intentionality, an issue arises as to the location of this intentional state. It may seem plausible to argue that collective intentionality entails the existence of a group mind (a ‘we-mind’) or a collective spirit that houses the collective intention. The fact that we-intentions are shared does not necessarily suggest that collectives have a unifying mind where the shared collective intentionality is actually located. \[16\] I propose that collective intentionality is located

---

16 Schmid here states that, “[c]ollective intentions, however, do not have a single subject. They have many. Thus the group mind is nothing we should be afraid of. It is
simultaneously in the minds (and the brains) of all the individuals that share the collective intentional state and that the interrelation of the minds and of the individuals is integral to the collective intentionality that the individuals share. Sharedness is a matter of relations between minds that transcends the physical boundaries of each individual, forming a network of minds and intentional states (Meijers 2003, 174). The sharedness and relationality of collective intentions is supported by the basic sense of collectivity that is common to all individuals; it permits the sharing of we-intentions. It is this sharedness that allows for social facts like money to exist and for the successful operation of our communities.

**COLLECTIVE INTENTIONALITY AND THE STATE**

The state theory of money uses the political authority of the state and the subsequent ability to levy taxes, to explain the emergence and persistence of fiat money. In order to complete the proposed ontological framework, it is necessary to show that state authority is not an ad-hoc explanation or a 'black box' without social foundations. It is therefore necessary to illustrate how it fits in the overall ontological framework of collective intentionality and constitutive declarations. The state enjoys the monopoly of power over a designated geographical territory as well as over the population of its subjects. The monopoly of power, taxation and fiat money indicates that the state is a very special institution; nevertheless, its authority and its sovereignty do not relieve the state from its dependence on society. The state is dependent on the collective intentions that constitute it—the shared representations about what the state is and does—and it is its status-function that organizes the relation to its subjects. State authority should be analyzed under the same ontological framework as money, comprised of collective intentions and constitutive declarations. The structure of governance is constituted on the basis of a fundamental intelligibility of political

merely a distorted individualistic image of a non-individualistic, holistic concept of the mind. Collective intentions are not intentions of the kind anybody has—not single individuals, and not some super-agent. For collective intentionality is not subjective. It is relational” (2003, 216).

17 Steinberger notes that “[w]hile particular states may differ [...] all states are essentially similar with respect to their proper scope of action, the nature of their authority, and their basic principle of organization. To demonstrate that this is in fact true and to show what it actually says about a state's activity, authority and internal constitution, is largely what it means to pursue an ontological theory of the state” (2004, 35).
action, i.e., a set of principles that define the expectations of the citizens in their relations to it and inform all aspects of social existence that have some relation to governance and regulation.\textsuperscript{18}

In \textit{Leviathan}, Thomas Hobbes defined the state in terms of sovereignty and order; since Hobbes, sovereignty remains the single most important defining characteristic of its identity (Hobbes 2012 [1651]). The status-function of the state is its ability to establish and maintain order over a specific geographical territory in virtue of its sovereignty in this geographical area. The monopoly over violence and the exclusive authority to regulate the use of force are consequences of its status-function as the sovereign enforcer of public order. The state can trump the actions of all other institutions and individuals precisely because of its monopoly over violence—this is what establishes its position in the social world making it the most powerful of all institutions. The special position of the state in the social hierarchy is already anticipated by its status-function as the sovereign institution: it is entrusted with maintaining order among all agents, including other institutions and organizations. The evolution of the state is a process of constant negotiation of the idea of state authority and its implications for governance. The state can expand (or limit) the scope of the exercise of its legitimate authority by constantly revising the content of the status-function upon which its legitimate authority is constituted. The stability and the adaptability of democratic societies lies in their capacity to constantly reinvent political purpose, to keep the authority of the government in check through regular elections that reinforce the collective intentions of the citizens in an evolving system of governance.

The collective intentionality of the citizens constitutes and regulates the state on the basis of the shared representations of what state authority and state power are, delineating the scope and the scale of its actions. Consequently, the efficacy of state actions and authority are

\textsuperscript{18} Steinberger further clarifies: “Indeed the state is nothing more than the authoritative manifestation of an entire way of life, reflecting as such, the full gamut of judgements about how things in the world—all things in the world—really are. It articulates and codifies a structure of truth about the nature of reality, i.e., the shared, typically tacit assumptions, presuppositions, theories, commitments and understandings on the basis of which individual members of the society are able to communicate intelligibly and interact coherently. Indeed, the propositions that constitute the idea of the state pertain not to this or that sector of society, but to the full range of social enterprises; it is composed of the notions of how institutional conflicts within the society are to be resolved for the good of society; it is a comprehensive structure of ideas that functions as a kind of rule-book of last resort, a final court of appeal on the basis of which all social disputes are evaluated [...]” (2004, 22).
conditioned by the common understanding of the identity of the state and the perceived legitimacy of its authority. As long as the actions of the state are an expression of its institutional status, the collective intentions of its citizens and the legitimacy of the state actions are safe. If state policies transcend the limits posed by the shared representation of the state and its functions, the state risks losing (some of) its authority which is constituted on collective intentionality. The state is an instantiation of the shared representation of how authority and power are organized and exercised; its legitimacy can be analyzed by the shared intentions that constitute its status as an authority.

Collective intentionality and sovereignty are in a relation of mutual dependence. The sovereignty of the state depends on the collective intentionality of the society at the same time as the state can use its sovereignty not just to inspire but also to enforce the collective intentionality, by the exercise of its power; this what Searle means when he says that, “power is a system of status-functions and thus rests on collective acceptance, but the collective acceptance, though not itself based on violence can continue to function only if there is a permanent threat of violence in the form of the military and the police” (2003, 204). The possibility to enforce collective intentionality through coercion undermines the proposition that the individual attitudes clustered collectively provide the source of all social facts and status-functions. The dissolution of individual autonomy is a consequence of the status-function of the state. The individual recognizes the state as the locus of sovereign authority that maintains order; consequently, the individual accepts his/her position as subject of the state’s power on the condition that the exercise of power is legitimate. Faced with the organized apparatus of political control and the monopoly of violence of the state, the individual may feel, and in fact be, powerless. The community may indeed be the source of all power; but for the state to maintain order it is necessary that each individual does not realize that the existence and authority of the state are dependent also on his or her collective intentionality (Searle 2003).

The same asymmetry of power between the individual and the state is characteristic of their economic and monetary relations. The credit relation between state and society, and the subsequent enforcement of the official currency is not voluntary, but is founded on the monopoly of the state and its ability to enforce taxation upon the citizens. Taxation is the consequence of the sovereignty of the state in the economic domain;
and with taxation emerges money and its acceptance by the public (Ingham 2004, 47-48). The necessity for the individual to earn income in the form of officially sanctioned currency in order to pay his/her taxes makes currency not only acceptable but desirable as a means of payment. The prerogative of the state to demand the payment of taxes and moreover to demand taxes in the tender that the state itself issues, lies in the center of the monetary system. Taxes cancel the debt that the issue of money creates: a loan of the issuing authority towards the bearers of money.

The source of economic sovereignty remains the product of collective intentionality of the public towards power and money; a collective intentionality that presupposes the acknowledgment not only of the monopoly of violence of the sovereign authority but also of its sovereign rights to sanction taxation and money. The state is also constrained by the attitudes of its subjects, i.e., by their expectations about the acceptability and value of money in the future. Tampering with the monetary system or imposing taxes beyond a point of considerable fairness by the public can lead to the loss of reliability of the state institutions; this can lead to increasing costs of enforcement, to inflation, to capital flight and the parallel circulation of other currencies. The attitudes of the public when expressed in concert can challenge the monetary sovereignty of the state in the same capacity that they support the existence and the circulation of money.

CONCLUSIONS
This paper outlined the ontological structure that underlies the emergence and persistence of fiat money based on the notions of constitutive declarations, collective intentionality, and sharedness, by illustrating the usefulness of these concepts for social analysis. The resulting theoretical framework is in many respects different from the usual treatment of money in collective intentionality literature (Hédoin 2013; Searle 1995, 2005; Smit, et al. 2011; Tieffenbach 2010). The underlying state theory of money is at odds with the concept of money as a means of exchange that emerges as the unintended consequence of the behavior of utility maximizing individuals. At the same time, the definition of collective intentionality as a relational and shared we-attitude that exists in interrelated individual minds contradicts the individualistic accounts of the exchange theory of money in analytic philosophy (Meijers 2003; Schmid 2003). The resulting ontology of
money can support the state theory of money, while it provides an ontological analysis of the state and its authority that is consistent with collective intentionality and constitutive declarations. The relation between state authority and the acceptability of money—a relation that carries the burden of explaining the emergence and persistence of money for the state theory of money—provides an important gap in the ontology of the theory.

REFERENCES


**Georgios Papadopoulos** has completed his PhD at the Erasmus University Rotterdam. Currently he is a visiting researcher at the department of aesthetics and communication at the Aarhus University and the University of Applied Arts in Vienna.

Contact e-mail: <languagegames@gmail.com>
Methodology in Capital in the twenty-first century: a “new-historical” approach to political economy

LUKE ANTHONY PETACH
Colorado State University

Abstract: This paper explores the methodological foundations of Thomas Piketty’s recent book Capital in the twenty-first century. The current literature on Piketty’s work lacks consensus as to which paradigm of economic thought Capital fits into (if any). In response to that literature, this paper argues that Piketty offers a new methodological direction for economic science in the form of an analytical ‘new-historicism’. The central emphasis of this methodology is an analysis of general dynamic laws on three levels: distribution, institutions, and history. A new-historical methodological framework applies new analytical tools to old economic problems raised by Smith, Ricardo, Marx, and others. This distinguishes Piketty’s framework from other contending paradigms or schools of economic thought, thereby alleviating confusion in the current literature surrounding Piketty’s book.

Keywords: Thomas Piketty, capital, economic methodology, history of thought, political economy

JEL Classification: B25, B41, E25

I. INTRODUCTION
Thomas Piketty’s (2014) Capital in the twenty-first century (from here on Capital) has been praised as “one of the watershed books in economic thinking” (Milanovic 2014, 519) and derided as a work whose conclusions “do not appear to be backed by the book’s own sources” (Giles 2014, 1). However, for all that has been said about Piketty’s book, it appears that more effort has been directed toward stating what Capital is not rather than what it is. Piketty’s work is not easily pigeonholed; it provides economists with an array of sweeping insights
backed by lengthy and comprehensive data sets. For its uniqueness, Piketty’s work should be praised. However, the uniqueness of Piketty’s methodology makes it difficult to classify his work. When trying to understand what exactly Capital’s methodology consists in, the expansiveness of Piketty’s analysis becomes problematic for those who wish to categorize the work within the history of economic thought. Institutionalist economists claim that Capital “ignores institutions” (Acemoglu and Robinson 2015, 1), Galbraith (2014a) critiques Piketty from a post-Keynesian perspective, Marxist economists do not support Piketty (Harvey 2014), while the Austrian school insists that Piketty is, in fact, a Marxist (Reisman 2014).

For all the disagreement about where Piketty’s methodology locates Capital in the spectrum of economic thought, there seems to be little agreement about what the methodological pillars of Capital actually are. But, before there can be agreement or disagreement about what school of thought Capital belongs to, there must be a comprehensive analysis of its methodology. This paper seeks to provide that analysis. In Section II this paper gives a general overview of Capital and identifies the central characteristics of the methodological framework. Section III responds to the lack of consensus in the literature, arguing that Piketty’s methodology offers a new direction for economic science in the form of a new-historical analysis. This is predicated on an interpretation of Capital that analyzes general, dynamic laws on three levels: the level of distributions, the level of institutions, and the level of history. Section IV identifies some sources of confusion about the fit of Capital in the history of economic thought. Section V offers some concluding remarks and suggestions for future inquiry.

II. CAPITAL IN THE TWENTY-FIRST CENTURY: AN OVERVIEW

The stated goal of Capital is a historical study of long-run distributional trends of wealth and income, rooted in “as complete and consistent a set of historical sources as possible” (Piketty 2014, 19). The text is structured according to three sections: an analysis of the capital/income ratio over time, an analysis of the structure of inequality (both wealth and income) over time, and policy recommendations based on those analyses. The historical analysis of Capital is woven into the framework of three “fundamental laws of capitalism”, the formation of which is guided by three central methodological pillars: distributional analysis, historicism, and analysis of institutions. The following paragraphs will
analyze the ways in which Piketty makes use of the three central methodological pillars of distributional analysis, historicism, and analysis of institutions in the context of the development of the three fundamental laws of capitalism.

First, it is necessary to discuss what exactly is meant by the term ‘law’. While the word—as it is used by Piketty—echoes language used by Marx and Ricardo, the laws that Piketty cites are not strictly empirical propositions (as in laws of nature). The first two fundamental laws are analytic propositions, a priori truths. The third fundamental law combines an empirical proposition with an analytic one. As will be noted later, the fact that \( r > g \) holds is not a historical necessity (there may be economies where \( r < g \)), but it is necessarily true that if \( r > g \), then inequality will increase.\(^1\) Given the nature of the fundamental laws, a better description might be to call them ‘principles’ (this fits with the tradition of the classical economists). Principles may be considered fundamental in the sense that they reflect relationships that influence the dynamics of all capitalist economies. In presuming that these relationships govern the dynamic properties of any (and all) capitalist economies, they have a law-like quality. For the sake of fidelity I will continue to use the term ‘law’ here, but it is used in the loose sense discussed above.

Piketty’s first fundamental law is a distributional law, it says that the share of capital in national income is a function of the product of the rate of return on capital and the capital/income ratio. In Piketty’s notation:

\[
\alpha = r \times \beta
\]  

(1)

Where \( \alpha = \) capital share in national income, \( r = \) rate of return on capital, and \( \beta = \) capital/income ratio.

This law, Piketty notes, “is a pure accounting identity. It can be applied to all societies in all periods of history, by definition” (2014, 52). Piketty’s first general law gives two key insights into his methodological

\(^1\) For those unfamiliar with Piketty, ‘\( r \)’ is the rate of return on capital and ‘\( g \)’ is the rate of growth of the economy. Piketty defines capital as “the sum total of nonhuman assets that can be owned and exchanged on some market” (Piketty 2014, 46). Growth is defined by Piketty as “the annual increase in income or output” (Piketty 2014, 25). It is a central contention of Capital that when \( r > g \) the distribution of wealth tends to become more unequal.
framework: the first concerns distributional analysis; the second concerns the formation of a general (or abstract) notion of capital.

The second methodological implication of Piketty’s first fundamental law has shown itself to be problematic. Some authors (Galbraith 2014a; Fullbrook 2014) claim that Piketty has either no concept of capital at all, or at best a very confused understanding of capital. In order to fully address these critiques and explain the methodological implications of Piketty’s first fundamental law, it is necessary to unpack the logic behind Piketty’s formation of an abstract notion of capital.

The classical political economists understood that to describe the dynamics of an economic system over time, it is necessary to have an abstract notion of how its parts function (so as to properly conceptualize the dynamic movement of the economic system). With the first fundamental law of capitalism, Piketty reveals the influence of the classical economists on his project. He writes: “[T]o summarize I define ‘national wealth’ or ‘national capital’ as the total market value of everything owned by the residents and government of a given country at a given point in time, provided that it can be traded on a market” (2014, 48). While Piketty’s notion of capital may not conform to the majority understanding of capital, this does not imply that he has no concept of capital at all (Fullbrook 2014). Although Piketty’s understanding of capital is more abstract than definitions of physical capital, it is nowhere near as abstract as others—such as the Marxian notion of capital as a social relation, or Bohm-Bawerk’s characterization of capital as a flow of dated labor quantities (Roncaglia 2005). In fact, Piketty’s measure of capital is essentially a financial valuation of different stock variables—something Galbraith (2014a) notes (which Piketty does not deny)—amounting to an equivalency with wealth. Wealth may be abstract, but it is for the most part measurable. In a later article Galbraith appears to soften his position, conceding “nothing prevents us from measuring \( r \) [the rental rate of capital]—as Piketty defines it—from the observed profit flow and financial valuation of the capital stock” (2014b, 146).

Whether or not Piketty’s understanding of capital is common, he argues that such an understanding is essential to engage in proper distributional analysis:

The capital/income ratio for the country as a whole tells us nothing about inequalities within the country. But \( \beta \) does measure the overall
importance of capital in a society, so analyzing this ratio is a necessary first step in the study of inequality (2014, 51).

Piketty’s first fundamental law of capitalism aptly demonstrates the first of the three central methodological pillars present in Capital, that of distributional analysis. It is not surprising that distribution would play a central role in Piketty’s methodology given his admiration for the classical project. For example, he states in Capital:

The economists of the nineteenth century deserve immense credit for placing the distributional question at the heart of economic analysis and for seeking to study long-term trends [...] It is long since passed the time when we should have put the question of inequality back at the center of economic analysis and begun asking questions first raised in the nineteenth century (2014, 16).

In turn, the second methodological pillar—evident in Piketty’s second fundamental law of capitalism—highlights the general historical framework, in which the entirety of his analysis takes place. This emphasizes the importance of taking into account historical contingencies during the process of economic analysis.

Unlike his first law, Piketty’s second fundamental law of capitalism is not an accounting identity (2014). Rather, as Milanovic (2014) points out, the second fundamental law of capitalism is a long-run equilibrium condition for the capital/income ratio in a given economy—Piketty calls it an asymptotic law (2014, 168). He expresses the law as follows:

\[
\beta = \frac{s}{g}
\]  

Where \( \beta \) = capital/income ratio, \( s \) = rate of savings, and \( g \) = rate of growth.

The long-run equilibrium condition does not address the historical factors that ultimately determine whether the capital/income ratio will be high or low (the third fundamental law does this), it simply tells what the long-run equilibrium capital/income ratio will be for a given rate of savings and growth. The establishment of long-run dynamic equilibrium conditions points to an important implication of Piketty’s methodology. In the context of Capital, proper economic analysis requires long time-horizons, especially in developed economies. Capitalist economies with
advanced financial markets may have extreme fluctuations in both
growth and savings patterns in the short-run, subsequently affecting the
distribution of income and wealth. In addition to fluctuations in asset
prices that occur in the short run, it is also possible for structural
adjustments or institutional shifts to occur that may hide long-run
trends. Piketty’s (2014) commentary on Simon Kuznets demonstrates
the problem with short-run frameworks.

Piketty challenges the conclusions that are drawn from Kuznets's
famous U-shaped curve, which shows the relationship between income
and inequality. Piketty contends that “[n]o generalized structural
process of inequality compression (and particularly wage inequality
compression) seems to have operated over the long run” (2014, 274).
Instead, it is the “budgetary and political shocks of two wars” (2014,
148) that proved to be the cause of the relationship observed by
Kuznets. As Piketty extends his analysis, the data he incorporates
appear to support his critique of Kuznets. The methodological
implication that Piketty draws from this—simultaneously reflected in
his second fundamental law—is that “a generation […] is the most
relevant timescale for evaluating change in the society we live in” (2014,
74). Accurate economic analysis must pursue a long-run historical
framework, or risk having insights obfuscated by short-run economic
fluctuations rooted in historical contingencies.

Piketty’s second law also reinforces the methodological principle of
distributional analysis introduced in the first law: “A country that saves
a lot and grows slowly will over the long run accumulate an enormous
stock of capital (relative to its income), which can in turn have a
significant effect on the social structure and distribution of wealth”
(2014, 166). Historical processes contribute to present distributional
arrangements, making it necessary to analyze the evolution of savings
and growth patterns over time in order to understand the present
distribution.

Piketty’s third and final fundamental law of capitalism—the famous
‘\( r > g \)’ inequality—represents a synthesis of the first two methodological
pillars of Capital; it is the clearest demonstration of Piketty’s third
methodological pillar, the importance of the analysis of institutions.

Piketty introduces the inequality relation ‘\( r > g \)’, citing it as “a
historical fact, not a logical necessity” (2014, 353). The third
fundamental law of capitalism states that so long as the inequality
relation ‘\( r > g \)’ occurs in an economy, “wealth originating in the past
automatically grows more rapidly, even without labor, than wealth stemming from work, which can be saved” (2014, 378). As a result, distributional inequality skyrockets as ownership of wealth becomes more and more unequal. However, that ‘\( r > g \)’ is regarded as historical fact and not logical necessity implies the following: “[I]ts truth depends, however, on the shocks to which capital is subject, as well as on what public polices and institutions are put in place to regulate the relationship between capital and labor” (Piketty 2014, 358). The methodological implication of interpreting ‘\( r > g \)’ as a contingent historical fact is that the distribution of income—and thus long-run historical trends, of which ‘\( r > g \)’ is one—can be altered by changes in public policy and institutional frameworks. Thus, institutions are key in determining economic outcomes. The prominent role of institutional analysis as a methodological tool is emphasized in Piketty’s (2014) discussion of the manner in which ‘\( r \)’ is itself determined. Without an analysis of institutions one cannot come to a complete understanding of the process by which ‘\( r \)’ is determined.

For Piketty the rental rate of capital is determined by the available technology and the valuation (or amount) of total available capital (2014). Colander criticizes Piketty’s model for “accepting marginal productivities [as given]” (Colander 2014, 162). This critique unfortunately misses the mark. Not only is Piketty critical of marginal productivity theory in general (a point to be revisited below), but the price of capital is only given insofar as the relevant technology and institutions remain unchanged. Piketty writes:

The price of capital, leaving aside the perennial short and medium-term bubbles and possible long-term structural divergences, is always in part a social and political construct: it reflects each society’s notion of property and depends on the many policies and institutions that regulate relations among different social groups, and especially between those who own capital and those who do not (2014, 188).

Piketty’s third fundamental law thus clearly highlights the importance of institutionalism as a methodological pillar in his analysis.

III. A NEW-HISTORICAL FRAMEWORK
Given Piketty’s use of distributional analysis, historicism, and institutionalism as the framework for his methodology, what can be said about where Capital falls in the history of economic thought? This paper
contends that Piketty (2014) offers—or at least hints at—a new-historical way of thinking about economics. In what way then, does Piketty’s work deserve this title?

First, the term ‘new-historical’ is intended to contrast with the moniker commonly applied to mainstream economics—i.e., ‘neoclassical’. It is because of the disregard for distributional issues that this comparison with neoclassical economics is necessary. The new-historical label serves a dual purpose with reference to its neoclassical counterpart. The first purpose is that it draws attention to Piketty’s methodology, viz. that it brings new insights to issues of historical import in the field of economics—e.g., Piketty’s article “Putting distribution back at the center of economics” (2015). The second purpose is to point out the ways in which neoclassical economics has (for the most part) ignored the important role played by class and distribution in the history of economic thought.

The second reason Piketty’s work deserves the title new-historical is because Capital marks a return to the tradition of economic thought—popular in both the classical political economy period and the pre-political economy period—that regards economics as both a scientific and a moral inquiry. Like many early economic thinkers, Piketty is concerned with the normative aspect of economic research. In some ways, the distributional analysis central to both the classical economists and Piketty is inherently normative. Piketty writes:

It is essential to carefully distinguish these various aspects and components of inequality [or of the distribution], first for normative and moral reasons (the justification of inequality is quite different for income from labor, from inherited wealth, and from differential returns on capital), and second, because the economic, social, and political mechanisms capable of explaining the observed evolutions are totally distinct (2014, 243).

The source of a given distribution of wealth—be it just or unjust—has critical implications for the response to that distribution, including whether or not that distribution should be altered through policy. One cannot ignore the ethical implications of an inequitable distribution; placing distributional analysis at the center of the new-historical framework requires one to put moral inquiry alongside it. New-historical economics spurns the logical positivism advocated by Friedman (1953). Economic science cannot afford to be ‘value-free’ if it wishes to ask valuable questions.
Finally, *Capital* represents a new-historical methodology because it is not merely historical. Piketty is not simply picking up where the classical economists left off. The new-historical classification is intended not only to distinguish Piketty’s (2014) work from the neoclassical paradigm, it is intended to show that Piketty’s methodology—while still maintaining the spirit of the classical mode of analysis—differentiates itself from the historical project of the classical economists in various ways.

Perhaps *Capital’s* most significant departure from the classical school is Piketty’s critique of marginal productivity theory—a theory originating in Ricardo’s corn model. While Piketty places himself in the classical tradition by developing a methodological framework of distributional analysis, historicism, and institutions, he diverges from the classical economists regarding the way in which the price of productive factors is determined relative to marginal productivity (and in turn, the way in which the price of those factors determines the distribution of income in an economy).

Piketty’s critique targets marginal productivity theory as it appears in neoclassical economics (one of the few theoretical tools of classical political economy to appear in neoclassical economics, albeit in an altered form); but, it is important to note that the theory finds its origins in Ricardo’s theory of differential rent, highlighting Piketty’s uniqueness. Piketty writes:

> The main problem with marginal productivity theory is quite simply that it fails to explain the diversity of wage distributions we observe in different countries at different times. In order to understand the dynamics of wage inequality, we must introduce other factors, such as the institutions and rules that govern the operation of the labor market in each society (2014, 308).

For Piketty, it is not enough to take the marginal output of the factors of production as given by the level of technology. To conclude that the distribution of income is determined by the prices of production that result from those marginal productivities is unrealistic. Additionally, not only is the marginal productivity of a factor of production determined by more than technology (e.g., it can be...

---

2 Although some scholars dispute whether or not Ricardo’s law of rent and the margin of production actually constitute the first instance of marginal productivity theory, at the very least one can say that it is here that the seeds for a future marginal productivity theory were sowed.
determined by human capital, availability of natural resources, etc.), but it is not marginal productivity alone that determines the prices of production. Piketty’s third methodological pillar reminds us that institutions must be accounted for, and different institutional arrangements can affect the distribution of income in ways that do not necessarily reflect the marginal productivity of various inputs. Syll highlights the intuitive logic behind Piketty’s rejection of marginal productivity theory: “put simply—highly paid workers and corporate managers are not always productive workers and corporate managers, and less highly paid workers and corporate managers are not always less productive” (Syll 2014, 40).

IV. RESPONSE TO THE LITERATURE

Much of the literature in response to Capital has taken a critical stance toward Piketty’s work (this is undoubtedly necessary and justified). However, the body of work that criticizes Piketty’s methodology—including the broader literature that seeks to evaluate Capital as a work from a particular school of thought—have been misguided. Once one views Capital as a new-historical work, many of the confusions found in the critical literature cease to exist. I have shown that the controversy surrounding Piketty’s definition of capital (also see Galbraith 2014a; Galbraith 2014b; Fullbrook 2014) becomes diffused when it is viewed in context, i.e., as an abstract notion of capital à la Ricardo or Marx. While there may be other examples in the literature, Acemoglu and Robinson (2015) and Colander (2014) provide two further cases where proper identification of Piketty’s methodological framework can clarify confusion.

Acemoglu and Robinson claim that “the question for general laws of capitalism is misguided because it ignores the key forces […] institutions and the political equilibrium” (2015, 1). Setting aside the emphasis that Piketty places on social and institutional factors in determining economic outcomes,3 Acemoglu and Robinson’s (2015) approach is flawed because it misinterprets Piketty’s ‘r > g’ inequality relation due to a lack of methodological context. Acemoglu and Robinson write, “‘r > g cannot be taken as a primitive on which to make

3 Note one particular portion of text here—Piketty writes: “the market is always embodied in specific institutions such as corporate hierarchies and compensation committees” (2014, 332). See pages 74-75, 140, 145-149, 188, 234, 237, 332, 356-358 of Capital for further discussion of the role of institutions.
future forecasts, as both the interest rate and the growth rate will adjust to changes in policy, technology and the capital stock” (2015, 11). The origin of this argument is unclear, because Piketty himself says just as much; recall his claim that ‘r > g’ is “a historical fact, not a logical necessity” (2014, 353). Acemoglu and Robinson (2015) seem to misinterpret the meaning of Piketty’s third fundamental law, due to a lack of proper understanding of the methodology of Capital. Once Piketty’s historicism and institutionalism are made obvious, it becomes clear that he does not take the third fundamental law to be “a historical primitive” but rather a historically contingent fact, the occurrence of which leads to inequality (but whose occurrence is in no way necessary). Endogenously determined interest and growth rates do little to undermine Piketty’s conclusions, given that the basis of the policy recommendations developed in the latter chapters of Capital depend on the premise that changes in institutions can prevent ‘r > g’ from occurring (i.e., the conclusion that r and g are in fact endogenously determined).

Colander deserves credit for being (mostly) correct in his analysis of Capital—he writes:

The problem with Piketty’s discussion is that it is based in a Ricardian, not a Millian, framework in thinking about the distribution problem [...] David Ricardo framed the income distribution question as a technical production issue. In Ricardo’s model technology determines marginal products and marginal products determine income distribution (2014, 161).

Colander’s analysis is correct insofar as Piketty’s analysis is Ricardian in origin. However, Colander misses that Piketty’s methodology is not merely classical, for Piketty’s critique of marginal productivity theory—insofar as wages and rents are not strictly determined by marginal productivity—represents a break from the Ricardian canon. By overlooking this differentiation, Colander finds himself arguing against a Ricardian notion that does exist in Piketty’s Capital.

V. CONCLUSION
Acemoglu and Robinson (2014), Colander (2014), and Galbraith (2014a; 2014b) offer three examples of interpretive confusion resulting from a lack of methodological classification of Piketty’s work. In this paper, I
have provided an analysis of the methodological foundations underlying Piketty’s (2014) work in an attempt to clear up this confusion. I have shown how a new-historical framework—based on distributional analysis, historicism, and institutionalism—is developed within the pages of Capital. I have examined the way in which the new-historical methodology differentiates itself from other contemporary paradigms of economic thought, and suggested ways in which acknowledging this differentiation leads to a better interpretation of Piketty’s work. Future research is needed to clarify and expound upon the principles of a new-historical system, and further, to address the ways in which a new-historical framework affects the mathematization of Piketty’s theory (see the problematic adaptation of Piketty’s (2014) theory to a Solow model by Acemoglu and Robinson (2015)).

REFERENCES

**Luke Anthony Petach** is a graduate student in economics at Colorado State University. His research interests include political economy, economic development, the history of economic thought, and problems concerning the intersection of philosophy and economics. Contact e-mail: <luke.petach@colostate.edu>
A welfarist critique of social choice theory: interpersonal comparisons in the theory of voting

AKI LEHTINEN
TINT, University of Helsinki

Abstract: This paper provides a philosophical critique of social choice theory insofar as it deals with the normative evaluation of voting and voting rules. I will argue that the very method of evaluating voting rules in terms of whether they satisfy various conditions is deeply problematic because introducing strategic behaviour leads to a violation of any condition that makes a difference between voting rules. I also argue that it is legitimate to make interpersonal comparisons of utilities in voting theory. Combining a realistic account of voters’ behaviour with a utilitarian evaluation of the outcomes then leads to the judgment that strategic voting is beneficial. If it is, then Arrow’s theorem does not have far-reaching consequences for democracy because one of its conditions is not normatively acceptable.

Keywords: economic methodology, strategic voting, interpersonal comparisons of utility, foundations of social choice

JEL Classification: B41, D72, D81

Insofar as social choice theory deals with voting, it aims to provide answers to two main questions. First, how well does voting perform in collecting information on individual preferences? Second, what is the relative performance of various voting rules? And which rule is the best?

The theory aims to answer these questions by imposing conditions on aggregation rules that are intended to represent the voting rules.

AUTHOR’S NOTE: Earlier versions of this paper have been presented at the public choice conference in Turku 2006, at a conference on theoretical and applied ethics in St. Petersburg 2013, and at the moral and political philosophy seminar at University of Helsinki in 2014. In addition to engaging me in fruitful conversations about the topics, Donald Saari has commented on an earlier unpublished version of this article (see Saari 2010). Eerik Lagerspetz and Hannu Nurmi have also made useful comments on various versions of the manuscript. Given that the aforementioned scholars may well disagree with some of my conclusions, it is clear that they could not be held responsible for the end result. Finally, two anonymous referees also helped me to make this a better paper.
The conditions are thought to be normatively compelling in the sense that they should be satisfied. They express how given inputs in terms of individual preferences are to be translated into collective judgments or decisions. For example, various monotonicity conditions specify that becoming more popular should increase the chances that a candidate is selected (or at least that the chances do not decrease); and the independence of irrelevant alternatives (IIA) condition requires, roughly, that the social choice between any two alternatives depends only on the individual preference orderings concerning those two alternatives.

The voting rules are then evaluated in terms of the set of conditions that the corresponding aggregation rules satisfy. The best aggregation rule is the one that satisfies the largest number of normatively important conditions. This is how the second question is answered. The famous theorems of Arrow (1951; 1963) and of Gibbard and Satterthwaite (Gibbard 1973; Satterthwaite 1975) provide one answer to the first question by showing that there are no aggregation rules that satisfy sets of conditions that are thought to be normatively compelling.

Social choice theory is only concerned with aggregating preferences or votes. In other words, it does not take behavioural assumptions into account, and voters are assumed to vote sincerely (i.e., non-strategically). In most voting rules, strategic voting means that a voter gives a vote to an alternative that she does not consider to be the most preferred one. One may consider the assumption of sincere behaviour as merely a consequence of division of labour among extant approaches: social choice theory evaluates voting procedures normatively by imposing conditions on aggregation rules, while various decision and game theoretical approaches study the positive properties of voting rules. Finally, implementation theory combines the normative and positive approaches. This division of labour seems sensible because strategic voting itself is widely thought to be self-evidently harmful, and thus unacceptable in a normative evaluation of voting.

However, I have recently shown that strategic behaviour increases the frequency with which the utilitarian winner is chosen compared to sincere behaviour in all commonly used voting rules studied thus far (Lehtinen 2007a; 2007b; 2008). The utilitarian winner is the candidate with the highest sum of utility for all voters. The basic reason for these results is that broadly supported candidates and alternatives are more likely to obtain strategic votes than those that have a number of vehement supporters, but also many enemies. To put it briefly,
these results show that strategic voting allows voters to express their preference intensities, and it is thus typically beneficial rather than harmful.

This paper endeavours to sort out the methodological, ethical, and normative consequences of my results for social choice and implementation theory. Although the argument is written in a self-contained way, some of the issues are already discussed in more detail in a previously published article (Lehtinen 2011). In that paper, I argued that even though an aggregation rule satisfies IIA, this does not mean that preference intensities or third alternatives do not affect the results, and it does not guarantee the observability of preference orderings or the absence of strategic voting. In this paper I generalise such argumentation to all central difference-making conditions. A difference-making condition is one that is satisfied by some, but not all, commonly used aggregation rules.

For the purposes of this paper, it is necessary to distinguish between an aggregation rule and a voting rule. The former refer to rules that operate on the basis of the assumption of sincere behaviour. The latter refer to voting rules in which strategic voting may affect the results.¹ My main claim can now be put as follows: while there are difference-making conditions for aggregation rules, there are no difference-making conditions for voting rules. Every voting rule violates every condition that is difference-making for the corresponding aggregation rule. If this claim is correct, it follows that comparing and evaluating voting rules in terms of difference-making conditions that the corresponding aggregation rules should satisfy is deeply problematic. This means that the conditions fail to provide reliable knowledge about the kind of information voting results really depend on.

Evaluating voting rules in terms of the conditions on the corresponding aggregation rules is the only way in which one can proceed in evaluating voting rules if interpersonal comparisons of utilities are not admissible. This is why I will argue against the view of some social choice theorists about the role of interpersonal

¹ There is thus a difference between what I call a voting rule and a voting procedure, as defined, for example by Blin and Satterthwaite (1977). The latter takes the vote ballots as arguments and yields social choices as output, while the former takes the utility profile as well as behavioural assumptions and beliefs as arguments, and yields social choices as output. A voting procedure does not make any behavioural assumptions because the procedure merely counts the votes irrespective of their relation to the underlying preferences of the voters. A voting procedure is thus an aggregation rule according to the terminology adopted in this paper.
comparisons in voting theory. In particular, according to this view, different voting (or aggregation) rules incorporate different interpersonal comparisons (or worse, that some but not all voting rules incorporate interpersonal comparisons). This view then leads to the idea that some particular voting rules are problematic because interpersonal comparisons are problematic. I will argue against this idea: Interpersonal comparisons rather concern the preference or utility profile, and thus they are completely independent of voting rules. This point is important for establishing the legitimacy of employing interpersonal utility comparisons in voting theory: it does not matter that it is impossible to obtain accurate information about interpersonal comparisons if the same comparison is made for every voting rule. At least voting theorists need not favour one rule over another by making interpersonal comparisons.

Insofar as social choice theorists refrain from making interpersonal comparisons, they never make normative judgments about the best alternative on the basis of the preference profile alone. All the normative work is done by the conditions imposed on the aggregation rules. My proposal for a more reasonable alternative for social choice theory is to define the best candidate independently of the voting rule, and then study voting rules in terms of how often and under what kind of circumstances the different rules find this candidate.

The paper is organised in the following way. I provide a synopsis of the overall argument in section 1. In section 2, I discuss the difference between votes and preferences. I then argue, in section 3, that the very idea that interpersonal comparisons are incorporated into an aggregation rule is not tenable. In Section 4, I make a case for a utilitarian evaluation of alternatives. Sections 2 to 4 thus describe how social choice theorists’ views concerning interpersonal comparisons arise in voting theory, and why they are flawed.

In section 5, I show how to generate a violation for any difference-making condition with strategic voting. This is done by considering examples from a recent debate between Donald Saari and Mathias Risse. In section 6, I explain the philosophical implications of the fact that strategic voting typically increases welfare. Sections 5 and 6 thus concentrate on evaluating aggregation rules in terms of normative conditions. In section 7, I discuss some constraints on the kinds of interpersonal comparisons that voting theorists could use.
1. **AGGREGATION RULES AND VOTING RULES**

Let \( p \) denote a profile of individual preferences; it assigns a preference ordering for each voter. Let \( U \) similarly denote a profile of individual utilities. An aggregation rule yields a social choice or a social ordering for a given preference or utility profile. For my purposes, it does not matter whether an aggregation rule yields an ordering, a single choice, or a set; and it also does not matter whether the rule takes the preference profile or the utility profile as its argument or whether the rule computes the outcomes from pairwise comparisons or orderings.

I will thus refer to preference and utility profiles simply as Input \( I \), and the values under aggregation rule \( A_j \) simply as Output \( O_j \). An *aggregation rule* can now be described as follows:

\[
O_j = A_j(I)
\]

Let \( \omega \) denote a theory of how voters behave when they vote. In particular, if voters are strategic, it specifies *how* they vote strategically when they do so. All realistic models of strategic voting assume that voters’ behaviour depends on their belief profile \( b \) as well as their utility profile \( U \) so that we can write \( \omega = \omega(b, U) \). Let \( V_j \) denote voting rule \( j \).

Since voters' behaviour also depends on the voting rule, we can write \( \omega_j(b, U) \). A *voting rule*, then, can be represented as follows.

\[
O_j = V_j(\omega_j(b, U)), \text{ or } O_j = V_j(\omega_j(b, I))
\]

The distinction between an aggregation rule \( A_j \) and the *corresponding voting rule* \( V_j \) is thus simply that only the latter takes strategic voting and other behavioural assumptions into account. Aggregation rule \( A_j \) corresponds to voting rule \( V_j \), and vice versa, if and only if the two types of rules process the incoming information in the same way. For example, the Borda aggregation rule corresponds to the Borda voting rule, and the Plurality aggregation rule corresponds to the Plurality voting rule, and so on. I will not provide specifications for \( \omega_j \) here. This was done in my earlier studies of strategic voting and in various papers in voting theory. The overall argument in this paper can now be put as follows:
1) Preference intensities are normatively important.

2) The best alternative is the utilitarian winner.

   [Social choice theory is based on the methodological injunction against making interpersonal comparisons. Thus, even if the utilitarian winner were to be admitted as the best alternative, if one has no way of observing which alternative is the utilitarian winner, it cannot be used in voting theory.]

3) However, interpersonal comparisons can be legitimately used in voting theory because the arguments against using interpersonal comparisons are not cogent:
   • Different voting rules do not incorporate different comparisons.
   • Interpersonal comparisons concern the utility profile rather than the relationship between the profile and the voting rule.
   • Although it is impossible to obtain reliable information about interpersonal comparisons, they can be made in a methodologically acceptable way in evaluating the performance of voting rules if the same comparison is made under every voting rule.
   • If one studies all voting rules with the same profile, one can make the same interpersonal comparisons under every rule.

4) [From 1, 2, and 3] The utilitarian criterion can and should be used in a normative evaluation of voting outcomes.

5) Strategic voting may well occur under incomplete information, and it does not require that voters have particularly precise information. Hence, it is not merely a logical possibility, but occurs reasonably often.

6) Every difference-making condition can be violated if voters are strategic.

7) Even if aggregation rule $A_i$ satisfies condition $C$, but aggregation rule $A_j$ does not, [from 6] there is no guarantee that $C$ is less frequently violated in voting rule $V_i$ than in voting rule $V_j$.

8) [From 7] Social choice theory fails to provide a satisfactory normative comparison of voting rules because the very idea of evaluating voting rules in terms of conditions on aggregation rules is misguided.

9) My models (in Lehtinen 2007a; 2007b; 2008) show that strategic voting is typically beneficial [this claim uses point 4].
10) [From 9] Social choice theory also fails to provide a satisfactory analysis of the first question because there cannot be an acceptable set of conditions that a voting rule should satisfy.

11) [From 8 and 10] The part of social choice theory dealing with voting is fundamentally flawed.

It is well known that Kenneth Arrow ignored the “game aspects” (1963, 7) of voting by assuming that people's preferences can be aggregated into a social choice or ordering without considering strategic voting. My critique could thus be taken to be disingenuous. However, taking those game aspects into account has more far-reaching epistemological consequences than has been previously acknowledged. My arguments for the steps above are all based on taking such consequences into account.

Steps 8, 10, and 11, are strong claims, and at this point might seem arbitrary, unclear, or unjustified. Some social choice theorists might attack 5 (i.e., that strategic voting does not require complete or perfect information) and then conclude that while 6 may be a logical possibility, 7 does not hold because strategic voting is too rare to be relevant. Then 8 would not hold either. In this paper, I will try to demonstrate 6, and then, taking 5 as given, argue for 7 and 8. Thus I will not try to argue for 5. It is primarily an empirical endeavour (see Mackie 2003), and it is not suitable to be discussed in a philosophical paper.²

If step 4 holds, then 9 is demonstrated in my earlier writings. The crucial steps are thus 1 (i.e., that preference intensities are normatively important), 2 (i.e., that the best alternative is the utilitarian winner), and especially 3 (i.e., that interpersonal comparisons can be legitimately used in voting theory). I presume that most social choice theorists acknowledge that 1 and 2 could be true, but they might claim that step 4 (i.e., that the utilitarian criterion can and should be used in a normative evaluation of voting outcomes) simply does not follow from 2. Even if the best alternative were the utilitarian winner, this information is useless if it cannot be identified. In particular, the sum of utility requires information on interpersonal utility comparisons but it is impossible to obtain reliable information on them. Given this epistemic problem, social choice theorists typically interpret voting

² For an explanation of why strategic voting does not require particularly precise information on the part of the voters, see Lehtinen 2007b. On how to measure the extent of strategic voting, see Kawai and Watanabe 2013, and the references therein.
rules as procedures for finding the best alternative. Rather than trying to define the best alternative in terms of the preference profile, the theory characterizes voting rules in terms of the conditions that the aggregation rules satisfy, and the conditions then define the best outcomes indirectly. Any outcome that could emerge from an aggregation rule that satisfies certain conditions is normatively acceptable in terms of satisfying those conditions. There is an aggregation rule that guarantees the choice of the utilitarian winner under the standard assumption that the ballots correspond to individual utilities: the zero-one rule (or range voting). However, this voting rule is so highly manipulable that very few social choice theorists believe that it could possibly be guaranteed to find the utilitarian winner.

On the one hand, various descriptive models of strategic voting (e.g., McKelvey and Ordeshook 1972; Enelow 1981) show how preference intensities affect the outcomes in many voting rules which seemingly only provide an opportunity to express preference orderings or mere top-preferences. On the other hand, as many authors have acknowledged, if we are only interested in defining the best alternative in terms of individual preferences, without regard to practical questions such as how to ascertain information about the socially most preferred alternative, preference intensities are obviously relevant (e.g., Dummett 1984, 54). I will thus not argue for steps 1 or 2. I take them to be obviously true. If step 1 holds, the only options seem to be the sum or the product of utilities. Intensities are thus acknowledged to be normatively and positively relevant, but if we are to believe social choice theorists, they should not be employed because interpersonal comparisons of intensities are meaningless (Arrow 1951; 1963) and epistemically suspect (Arrow 1977).³

What needs to be established is thus not that preference intensities are normatively or positively relevant (steps 1 and 5), but rather why the arguments against interpersonal comparisons fail to support the conclusion that a normative evaluation of voting rules is best conducted without making interpersonal comparisons. In this paper, I do not distinguish between interpersonal comparisons of preference intensities and of utilities, and I always mean what is referred to as ‘full comparability’ in the literature on social welfare functionals (SWFLs). Many social choice theorists think that, in contrast to intensities, it is

³ Arrow (1978) acknowledged the possibility of ordinal interpersonal comparisons.
It is possible to obtain reliable information on preference orderings, and that different voting rules make different interpersonal comparisons.

I will argue that the arguments against making interpersonal comparisons of intensities would be cogent if the assumption of sincere behaviour characterized voting. But it does not. I will thus provide an argument for making interpersonal comparisons of preference intensities in voting theory. Since making interpersonal comparisons is necessary only in a normative evaluation of voting rules, but not in predicting voters’ choices, this is tantamount to providing an argument for evaluating the candidates or alternatives in utilitarian terms. However, although I am defending a utilitarian evaluation of outcomes in evaluating voting rules, I do not intend to defend utilitarianism in general. Neither can I be defending or criticising any particular voting rules, especially the utilitarian ones, because, as I will show, interpersonal comparisons concern the utility profile.

If voting theorists are not allowed to make interpersonal comparisons, they are not allowed to pass normative judgments on the alternatives directly on the basis of the utility (or preference) profile.\(^4\) Once this requirement is adopted, examining the conditions that the various aggregation rules satisfy (and the paradoxes associated with violating them) becomes the only way in which voting rules can be normatively evaluated. It has long been acknowledged that the normative force and importance of various conditions is a matter of controversy, and the problem of weighing the importance of different conditions easily explains why social choice theorists do not agree on which voting rules are the best.

There is a much more serious problem, however, with the very idea of evaluating voting rules in terms of conditions on aggregation rules. Although there are differences between the sets of conditions that the various aggregation rules satisfy, due to strategic voting, the corresponding voting rules never satisfy any conditions that make a difference to their comparative evaluation. It is thus impossible to make meaningful comparative normative assessments on voting rules with such conditions, because we ought to be interested in whether voting rules, rather than aggregation rules, satisfy various conditions. This argument establishes that eschewing interpersonal comparisons forces

\(^4\) As mentioned at the beginning of this section, an ordinal preference profile is a list of preference orderings, one for each voter. A utility profile is a listing of all utilities in an electorate, one utility number for each individual and each alternative.
upon us a highly dissatisfactory method of comparative normative evaluation.

Step 10 (i.e., that there cannot be an acceptable set of conditions that a voting rule should satisfy) can be explicated as follows. We know that strategic voting cannot be prevented in any real voting rule (Gibbard 1973; Satterthwaite 1975). If strategic voting is beneficial, there cannot be a set of normatively appealing conditions that any voting rule could ever satisfy, because strategic voting implies that all other difference-making conditions will be violated. It follows that the set of conditions for Arrow’s theorem are normatively unacceptable, or, to put it more strongly, that there cannot be a set of normatively acceptable conditions in the first place.

2. DISTINGUISHING BETWEEN PREFERENCES AND VOTES

When Kenneth Arrow founded the modern version of social choice theory in the 1950s, only ordinal utilities were considered epistemically acceptable. Interpersonal comparisons of utilities were considered even more problematic than cardinal utilities because there are no individual choices that could provide reliable information on such comparisons. Myerson (1985) gives an often cited expression of such epistemic arguments in pointing out that it is not possible to choose to be person A in state Y rather than person B in state X. Interpersonal comparisons are epistemically problematic because it is impossible to obtain choice-based information which is similar to information obtained from an elicitation procedure such as the reference lottery technique (see, e.g., Hirshleifer and Riley 1992), and which would allow definition of interpersonal differences and similarities in utilities. Roy Harrod links this epistemic argument with the scientific respectability of economics (Harrod 1938, 395-396; see also Robbins 1938; 1952).

To motivate the discussion, consider some recent views about interpersonal comparisons. Mathias Risse argues as follows:

Often, for instance, we would like to be able to use information about the strength of preferences, or use more fine-grained judgments, say, as provided by a point system. Yet what matters for the choice of an aggregation rule is not merely what information should be used given the purpose of the decision process, but also what information one can reliably elicit. There are two concerns. On the one hand, the more fine-grained information we admit, the more
interpersonal comparability among individuals is problematic […] The other problem is manipulability (Risse 2009a, 797; see also Risse 2004, 58).

Laslier (2012) reports the results of a poll conducted at a meeting of voting theory specialists concerning what they think about the performance of various voting rules. The participants were given the chance of providing their reasons for the choices they made. Here is one comment:

The Condorcet Principle takes into account only the ordinal preferences of every voter between any pair of alternatives because attempting to take into account also voters’ cardinal preferences (as under the Range Voting procedure) would not only imply that a Condorcet winner⁵ may not be elected or, worse, that a Condorcet loser may be elected, but also that inter-personal comparisons of utility are possible and acceptable—which they are not! (Dan Felsenthal, in Laslier 2012, 339).

Risse and Felsenthal are thus arguing that some voting rules incorporate interpersonal comparisons, particularly ones that aim to collect information on preference intensity, and some others do not.⁶ I will now argue that such a view is based on a rather elementary confusion between voters’ underlying preferences and expressed votes.⁷ I will proceed by discussing an example that Arrow used in arguing for the IIA condition. My concern here is not so much with IIA, but rather with showing how the failure to distinguish between vote ballots and preferences leads to the mistaken view that some but not all voting rules incorporate interpersonal comparisons.

The assumption of sincere voting makes it impossible to formally distinguish between preferences and vote ballots within the theory. The theory thus proceeds on the assumption that preferences directly affect the outcomes of voting rules. This, in turn, leads to the idea

---

⁵ A Condorcet winner is a candidate that has a majority against each of the other candidates, so that it beats any other candidate in a pairwise contest.

⁶ Risse and Felsenthal do not represent the views of all social choice theorists. Saari reports the following in a footnote: “Some readers of earlier drafts of this paper claim that the Borda Count has nothing to do with interpersonal comparisons, others claim it does” (Saari 1998, 258).

⁷ It is not that votes and preferences are never distinguished in social choice theory. Saari (2003) says that the former describe what happens with a particular rule, not what should happen based on the voters’ preferences. See also Reynolds and Paris 1979; Austen-Smith and Banks 1999.
that different voting rules collect different information on preferences, and that the differences have to do with how voters may express their preferences. Thus the idea arises that different voting rules make different interpersonal comparisons. I will argue that this assumption is misguided because strategic voting implies that we cannot assume that the votes correspond to preferences. When we compare the outcomes of voting rules, insofar as the outcomes are evaluated by employing interpersonal comparisons, the comparisons are independent of the voting rule. By the end of section 4 below, I will come to the conclusion that interpersonal comparisons are methodologically legitimate if they concern the utility profile because the same profile is used to study different voting rules.

Arrow introduced IIA in order to restrict social-welfare judgements to those that may be the result of voting processes (e.g., Arrow 1973). Insofar as an aggregation rule is assumed to represent a voting rule, this is tantamount to identifying such judgements with the outcomes of voting rules. Arrow’s theorem is thus concerned with evaluating the performance of voting rules.

Arrow’s discussion concerns the so-called zero-one rule. Each person’s utility of the most preferred alternative is normalised to one and the least preferred alternative to zero. The rule may be interpreted as a version of utilitarianism understood as an ethical doctrine, as well as a decision-making procedure. Interpreting the zero-one rule as a decision-making mechanism means that if there are three alternatives each voter ‘i’ will report his or her preference ordering by assigning one point to the best, zero points to the worst, and \( v_i \in (0, 1) \) to the middle alternative. The outcome is determined by adding these points. The alternative with the largest number of points is selected. If there are only two alternatives, voters assign one point to the better one and zero points to the worse.

Arrow’s example is shown in Table 1 below. Three individuals: 1, 2, and 3, have the following preferences for alternatives A, B, and C:

---

8 Isbell (1959) was the first to formalise this rule and to suggest that it could be used for interpersonal comparisons. Dhillon and Mertens (1999) provide an axiomatic justification for it. See also Sheng 1987; and Hausman 1995.

9 If it is interpreted as a decision-making mechanism, its functioning is equivalent to “utilitarian voting” (Hillinger 2005) and “range voting” (Smith 2000). Tangian (2000) calls it “Laplace’s method”.

---
Arrow states that this zero-one assignment of utilities is “designed to make individual utilities interpersonally comparable” (1963, 32). He thus views the numbers in parentheses both as points that voters may express and as interpersonally comparable utilities.

Arrow presents two related arguments against this rule for making social choices. They are also arguments for IIA, and both are based on a modification of voter preferences. According to the first, deleting the worst alternative, C, from the set of available alternatives should not affect the final outcome. The utilities would be as in Table 2.

Table 2: Arrow’s example with C deleted

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(1)</td>
<td>A (1)</td>
<td>B (1)</td>
</tr>
<tr>
<td>B</td>
<td>(0)</td>
<td>B (0)</td>
<td>A (0)</td>
</tr>
</tbody>
</table>

Under the second modification, the first two individuals are indifferent between A and C, and the third now finds C indifferent to B, while the relative positions of A and B are unchanged in all individual orderings. “Then the assignment of utilities to A and B becomes the same as it became in the case of blotting out C entirely, so that again the choice between A and B is altered, contrary to Condition 3 [IIA]” (Arrow 1963, 32). The utility assignments after these changes are shown in Table 3.

Table 3: Arrow’s example with changes in preferences

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A C</td>
<td>(1)</td>
<td>A C</td>
<td>B C</td>
</tr>
<tr>
<td>B</td>
<td>(0)</td>
<td>B (0)</td>
<td>A (0)</td>
</tr>
</tbody>
</table>

IIA is violated because ordinal preferences between A and B remain unchanged, but social choice changes from B in Table 1 to A in Table 2,
and the social ranking between A and B changes between Tables 1 and 3. In Arrow's account, the utilities for the alternatives are first as follows:

A: \(1 + 1 + 0.5 = 2.5\)
B: \(0.9 + 0.9 + 1 = 2.8\)
C: \(0 + 0 + 0 = 0\).

Blotting C out of existence then changes them to:

A: \(1 + 1 + 0 = 2\)
B: \(0 + 0 + 1 = 1\).

Changing the relative position of C changes them to:

A: \(1 + 1 + 0 = 2\)
B: \(0 + 0 + 1 = 1\)
C: \(1 + 1 + 1 = 3\).

Here is how Arrow makes his point:

If C were blotted out of existence, it should not make any difference to the final outcome; yet, under the proposed rule for assigning utilities to alternatives, doing so would cause the first two individuals to have utility 1 for A and 0 for B, while the third individual has utility 0 for A and 1 for B, so that the ordering by the sum of utilities would cause A to be preferred to B (Arrow 1963, 32).

It is customary to think that IIA is mainly concerned with ruling out information on preference intensities (or cardinal utility). However, it is clear that it “does not rule out the existence of preference intensities […] The construction of these relative preference intensities may involve irrelevant alternatives, but IIA is still satisfied” (Hammond 1987, 200). Hammond’s point is that it may be possible to find out about intensities for the various alternatives only by using information on “irrelevant” alternatives, but this has nothing to do with the existence of such intensities. Hammond suggests that preference intensities underlie the choices of voters (see also Mongin and d'Aspremont 1998; Ng 2000, 68). It is clear that such an interpretation treats intensities as existing independently of the voting rule, and that the latter is not interpreted as a preference-elicitation procedure. It is instructive to think of intensity as referring to some substantial notion of utility such as desire.
satisfaction or feelings of pleasure. It seems to make sense to say that the substantive utilities of voters would be different under different outcomes, and that they are independent of their choices in the counterfactual sense that a voter's substantive utility could be assumed to be the same if A (or B or C) was selected, irrespective of whether he or she voted for it.

If 'utility' refers to something substantive, the numbers we use to describe voter preferences remain fixed irrespective of the choices that they may be able to make. In contrast, according to the modern conception of utility (such as von Neumann and Morgenstern's), utility numbers represent a person's preferences or choices, and are not connected to any substantive notions. Ultimately, the utility numbers are only based on choices because they are supposedly constructed in choice experiments such as the reference lottery. It is consistent with this viewpoint to say that voters 1 and 2 have utility zero for outcome B in Table 2, even though they had utility 0.9 in Table 1. (Since the utility numbers are not unique, however, one could just as well use any two numbers such that the number for A is larger than that for B.) However, if the utility number 0.9 refers to substantive preference intensity, it makes no sense to say that it disappears when we move from Table 1 to Table 2: it is just that voters are no longer able to express this preference intensity by directly reporting a cardinally measurable number.

If a voter gives her vote to B when A and C are available, we cannot even say, merely on the basis of this choice, that she prefers B to A and C, because she may have voted strategically. If such a choice were taken to define her preferences, it would define them falsely. Whatever plausibility revealed-preference arguments have elsewhere, in the voting context they are hardly compelling, and the whole idea of utility as a representation of preference cannot be meaningfully applied. I do not intend to claim that social choice theory is committed to revealed preferences. My point here is rather that if one were to appeal to such an idea so as to defend the identification of preferences and votes, it would be a highly unsatisfactory defence.

Arrow analyses the example as if there were first an intensity of preference between A and B which, however, suddenly disappears when C is deleted. It is obvious, however, that he did not intend to claim that the underlying preference intensities had actually changed. Had they done so, there would have been no reason to worry about the change in
the outcome. If the utilities in Tables 1, 2, and 3 are interpreted as substantive preference intensities, a utilitarian would accept the change in outcome from B to A, precisely because they change. It is true that, if we take Arrow literally, he seems to be saying that cardinal preferences between A and B change while ordinal preferences remain the same. However, this would be misreading his obvious intention. His point is clearly to show that this method of voting yields unsatisfactory outcomes, because voters are first able to express cardinal preferences for three alternatives, but after the change they may only express their ordinal preferences for two alternatives.

Let us now see how the example should be analysed if it is assumed that the utilities in Table 1 represent intensities of desire satisfaction that underlie voters’ choices. Table 4 displays the preference intensities for the two remaining alternatives after the elimination of C.

**Table 4: Arrow’s example modified**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(1)</td>
<td>A</td>
<td>B (1)</td>
</tr>
<tr>
<td>B</td>
<td>(0.9)</td>
<td>B (0.9)</td>
<td>A (0.5)</td>
</tr>
</tbody>
</table>

Table 5 displays similar intensities following the second modification.

**Table 5: Arrow’s example modified and remodelled**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C (1)</td>
<td>A C (1)</td>
<td>B C (1)</td>
</tr>
<tr>
<td>B</td>
<td>(0.9)</td>
<td>B (0.9)</td>
<td>A (0.5)</td>
</tr>
</tbody>
</table>

These tables show that B is more preferred than A in utilitarian terms irrespective of whether C is included or not in the set of alternatives. It is an altogether different matter that voters would be able to express only their ordinal preference if there were only two alternatives. A utilitarian would then say that B should be selected but that A is selected.

Arrow’s example does prove what it was intended to prove, viz., that the zero-one rule violates IIA if it is interpreted as a decision-making
mechanism. The problem with this analysis is rather that Arrow seems to assume that, since voters may express points with a cardinal significance under the zero-one rule, these choices simultaneously provide information about their *interpersonally comparable utilities*.

### 3. Can Interpersonal Comparisons of Utility Be Incorporated into Voting Rules?

Social Welfare Functionals (SWFLs) are the standard tool for investigating interpersonal comparisons in social choice theory (see, e.g., d’Aspremont and Gevers 2001; Bossert and Weymark 2004; Blackorby and Bossert 2006). In Sen’s proof of the impossibility theorem with cardinal utility functions (1970, 128-130), non-comparability of utility means that the ordering provided by the social welfare functional is the same irrespective of which linear transformation of the utility profile is used. This means that even though the SWFL approach discusses interpersonal comparability, no judgement is ever formulated on the candidates (or the choice alternatives) in terms of a *given* utility or preference profile alone, but rather only insofar as intensities or interpersonal comparisons affect the SWFL. The essence of Sen’s proof is that, since the SWFL yields the same ordering irrespective of which linear transformation of the utility profile is used, cardinal utilities provide the same information for it as ordinal utilities. It follows that cardinal utilities do not provide any information additional to that derived from preference orderings. An impossibility result can then be proven with cardinal utilities, because it can be proven with ordinal ones.

Hammond and Fleurbaey (2004) criticise this approach for its failure to provide an unambiguous procedure for *embodying* or *incorporating* the interpersonal comparisons into the SWFL. I will now argue, in the spirit of Hammond and Fleurbaey, that voting rules cannot incorporate interpersonal comparisons even in principle. In other words, voting rules can never be represented by SWFLs. Incorporating interpersonal comparisons in a voting rule could be taken to mean one of two things: that people may somehow express interpersonal comparisons when they vote, or that when the outcomes of a voting rule are evaluated

---

10 This claim must be qualified slightly. Bordes and Tideman (1991) argue that IIA is defined for a situation in which the set of alternatives is fixed and where the preference profile changes. Arrow’s first modification thus does not qualify as a violation of IIA because it is undefined, but the second one does.
normatively, it is necessary to make interpersonal comparisons. I will show that the first interpretation is false, and that while the second interpretation is sensible if the normative evaluation is to be based on the utility profile, the interpersonal comparisons are not different under different voting rules.

Voting choices cannot express interpersonal comparisons. The argument for this claim is quite simple. It follows from Myerson's argument that it is not possible to choose to have alternative A selected in a voting rule with i's preferences rather than alternative B with j's preferences.\(^{11}\) Hence, interpersonal comparisons cannot be expressed in terms of individual voting choices, even in principle (see Sen 1999, 355), because each individual's choices only provide information on his or her preferences, but cannot yield any interpersonal information.

Arrow's initial idea was presumably that the zero-one rule violates IIA and incorporates interpersonal comparisons, but that the so-called method of majority decision does not incorporate interpersonal comparisons and does satisfy IIA.\(^{12}\) This method is defined as follows. If a majority of voters prefer alternative A to B, then A is higher in the social-preference relation than B. Clifford Hildreth argued that the method of majority decision involves an interpersonal comparison "because it chooses between A and B by comparing the number who would gain utility in passing from A to B with the number who would lose" (Hildreth 1953, 91). He concluded that the issue was not whether we could do without interpersonal comparisons, but rather what sorts of comparisons we would be willing to make. Hildreth thus argued that all voting rules involved them, and that different rules incorporated different comparisons.

Different voting rules could be taken to incorporate different interpersonal comparisons if the utilities expressed by voters could be aggregated by assuming that the method of adding points or scores would incorporate the relevant comparability assumption. The zero-one rule, conceived of as a decision-making rule, would thus incorporate

---

\(^{11}\) Even though I apply Myerson's argument here, I do not wish to deny the possibility of making such comparisons by claiming that they cannot be elicited. After all, everybody—including Lionel Robbins (1981)—admits that we make interpersonal comparisons in our everyday lives. Furthermore, our knowledge of interpersonal comparisons is in fact based on choices (Little 1957; see also List 2003). It is just that these choices are more difficult to interpret than those made in a reference lottery experiment.

\(^{12}\) He never states this explicitly, but it seems to me to be a fair interpretation of his intention.
interpersonal comparisons of intensities if the points were identical with the preference intensities. However, the example discussed above shows that we have reason to believe that this is often not the case if intensities are interpreted in a substantive way.

Hildreth seems to be arguing that, since voters can only express their preference orderings (or rather, pairwise preferences) under the majority-decision method, but can also express preference intensities under the zero-one rule, then utilities are cardinally comparable in the latter and ordinally comparable in the former. And indeed, if one assumes that the vote ballots correspond to utilities, then the zero-one rule incorporates interpersonal comparability of preference intensities because voters are allowed to express their *intrapersonal* intensities “directly”, that is, without misrepresenting their preferences.

We only have knowledge on the expressed points in the voters’ ballots. But then, given that the vote ballots are supposed to reflect the underlying preferences, social choice theorists assume that they correspond to one another. Hence they arrive at the judgment that the zero-one rule is problematic as concerns interpersonal comparisons because the theorist must assume that the preference intensities are interpersonally compared when the voting rule computes their sum: the weight of each individual voter is assumed to be the same.

However, voters do not express their interpersonal or even intrapersonal intensities of preference directly because they only report *points* or *numbers* that may or may not have a one-to-one relation to them. The problem with Hildreth’s argument is thus that it blurs the distinction between the kind of information voters are able to express sincerely with their votes, and the kind of interpersonal comparisons that are required for interpreting or evaluating the outcomes under various voting rules. It may be natural to conclude that different voting rules come with different comparability assumptions if one adopts the standard social-choice assumption that the votes or points correspond to individual utilities. After all, if the points do correspond one-to-one to voters’ utilities, it seems as if the rule is making an interpersonal comparison because it is adding points across individuals. Furthermore, given that there is no other way of obtaining information on interpersonal comparisons of utilities in the voting context, some social choice theorists have assumed that the rules are making the comparisons.
The main problem with this view is that if the rules were to be making the comparisons, they would be wildly inaccurate in every rule rather than just in the zero-one rule. If strategic voting is taken into account, every voting rule is likely to reflect at least some intensity information at the aggregate level. Hildreth’s conclusion that different voting rules make different interpersonal comparisons is false if the expressed votes do not necessarily correspond to their utilities.

I will now show that the votes do not necessarily correspond to utilities. Consider how voters might act under the zero-one rule. Let us assume—applying the logic of Ordeshook and McKelvey (1972)—that voters rank the candidates in terms of expected gains. Let $p_{jk}$ denote the probability of being pivotal between candidate $j$ and $k$, and let $U_j$ denote the utility for candidate $j$. The expected gains for $j$ are given by $E_j = \sum_{j \neq k} p_{jk}(U_j - U_k)$. It seems safe to assume that in Arrow’s example, the pivot probabilities $p_{AC}$ and $p_{BC}$ are equal to zero because $C$ never has a chance to win. For voters 1 and 2, the expected gains are then $E_A = p_{AB}(U_A - U_B)$ and $E_B = p_{AB}(U_B - U_A)$. As long as $p_{AB}$ is positive, $E_A$ is positive but $E_B$ is negative. If voters give positive points only to those candidates for whom the expected gains are positive, they might end up giving a bullet vote to $A$ so that the vector of points is $(1, 0, 0)$ for $(A, B, C)$. Bullet voting means giving maximum points to one candidate and zero to the rest even though the voting rule allows for expressing more fine-grained intensity information. If this is how voters behave, the outcomes from the zero-one rule end up depending on the same kind of information that the plurality rule would take as input under the assumption that voters are sincere.

Voting rules such as range voting or the zero-one rule, which might yield precise information on preference intensities under sincere behaviour, may end up providing only top-preference information under strategic behaviour, and voting rules such as plurality rule that only collect top-preference information under sincere behaviour may end up providing rich intensity information under strategic behaviour.\footnote{There are other voting rules that allow voters to provide at least some intensity information such as Approval voting (Brams and Fishburn 1983) and Majority judgment (Balinski and Laraki 2007; 2010). I have shown (Lehtinen 2010) that there can be bullet voting in the former which is problematic under behavioural heterogeneity. I have thus far not been able to construct a model of strategic voting under majority judgment, and I have not been able to study whether there could be bullet voting in this rule. Whether its relative resistance to manipulation is an argument in its favour or against it is thus an open question.} Even though each individual voter’s choice only indicates ordinal preferences
in many voting rules (except under the aforementioned forms of utilitarian voting), preference intensities affect the results in all voting rules. This, in turn, follows from the fact that the conditions for strategic voting always depend on preference intensities (e.g., McKelvey and Ordeshook 1972; Enelow 1981). Thus, even though individuals provide their information in ordinal form, and even though the voting rule provides only a ranking or a single winner, the outcomes do depend on preference intensities through strategic voting. No voting rule is thus able to only depend on ordinal information. If the kind of information voters are able to express with their strategic votes were to determine the kind of interpersonal comparisons that a voting rule makes, this would mean that every voting rule makes interpersonal comparisons of preference intensities. The absurdity of this claim should make it easier to understand that the kind of information that voters are able to express, or the form in which vote information enters or exits the voting rule, have nothing to do with interpersonal comparisons of utilities.

It follows that one cannot reasonably define interpersonal comparisons on the basis of the kind of information that voters are allowed to express, or on the basis of the kind of information that ends up affecting the results in any voting rule. If voting theorists make interpersonal comparisons, they do so in order to be able to interpret voting outcomes in welfarist terms. We have seen that there is no reason to think that they are different under different voting rules, merely because voters may provide different information about their preferences sincerely under different rules. Interpersonal comparisons thus have nothing to do with the various voting rules, and different voting rules do not incorporate different interpersonal comparisons.

It follows that neither IIA, nor any other condition could rule out interpersonal comparisons because it would be a category mistake to claim that it does: IIA concerns the relationship between individual preferences and the outcomes of aggregation rules but interpersonal comparisons have nothing to do with the latter.

What, then, do interpersonal comparisons concern if they cannot concern the voting rule or the relationship between the utility profile and the voting rule? I take it to be obvious that they must concern the utility profile itself. Now that we have put the object of interpersonal comparisons in its proper place, let us consider why the impossibility of obtaining reliable information on interpersonal comparisons does not
provide an insuperable obstacle to using them in evaluating voting rules normatively.

4. WHAT FOLLOWS FROM THE EPISTEMIC ARGUMENT AGAINST INTERPERSONAL COMPARISONS OF UTILITY?

As we have seen, the epistemic argument (Myerson 1985) against interpersonal comparisons posits that there is no choice-based procedure for obtaining information on interpersonal comparisons. Furthermore, it is often argued that it is more difficult to obtain information on preference intensities than on orderings, and I presume that this is the reason why social choice theorists consider Risse and Felsenthal’s view plausible. Furthermore, the literature on social welfare functionals has studied various comparability assumptions and shown how different social choice functionals require different comparability assumptions. The rationale for studying SWFLs—in other words different notions of comparability—then hinges on there being such epistemic differences between the various aggregation rules.

Just like many other writers on interpersonal comparability, I do not think that it is possible to obtain reliable information on interpersonal comparisons of preference intensities, and that interpersonal comparisons of intensities are epistemically more difficult than, say, comparisons of levels of utility. However, I will now argue that the injunction against making interpersonal comparisons (of intensities) in evaluating voting rules simply does not follow from the epistemic argument.

We could interpret Arrow’s treatment of the zero-one rule by suggesting that the social choice theorist makes the interpersonal comparison once she assumes that the points correspond to voters’ utilities. One way of making sense of the idea that some voting rules require more precise information on interpersonal comparisons than others would be to argue that in order to use the zero-one rule, we would have to obtain information on interpersonal comparisons of preference intensities. But since the zero-one rule is highly manipulable, the rule does not provide reliable information about preference intensities. This is the basis for Risse’s claim—I believe—that we cannot obtain reliable information on preference intensities. I agree with Risse that trying to elicit information on preference intensities with rules like range voting or the zero-one rule may be very difficult. But the reason for this is not that such rules may be more susceptible to manipulation
than other rules. I believe they are, but that is not the main shortcoming of such rules. It is rather that strategic voting leads to bullet voting, and such voting rules may thus end up giving us less intensity information than other rules which ostensibly only collect information on preference orderings or top preferences!

Arrow thought that IIA guarantees the observability of preference orderings, and indeed it does this in his original framework because it assumes sincere behaviour. Thus, there would be a case against preference intensities and interpersonal comparisons if people never voted strategically. Note that the majority rule is not manipulable if it is used to choose from two alternatives. However, in order to apply it (or “method of majority decision”) when there are more than two alternatives, one has to specify how exactly one is to make all the pairwise comparisons, and how exactly one is to use the information from such comparisons. This is why the majority rule translates into a large number of aggregation rules when there are more than two alternatives (e.g., Borda, Copeland, amendment agendas, Kemeny). I have shown (see Lehtinen 2011) that the outcomes depend on preference intensities under amendment agendas. Such agendas are a common way of implementing majoritarian voting, and they satisfy IIA. This means, however, that if voters act strategically, satisfying IIA does not mean that a voting rule precludes preference intensity information or that satisfying IIA guarantees observability of preference orderings. Moreover, IIA itself does not preclude strategic voting (see Saari 2008, 60; and Lehtinen 2011 for a detailed argument). Risse thus proceeds on the false presupposition that majoritarian voting rules are not affected by preference intensities. Although these rules do not explicitly ask voters about their individual intensities, they do affect the outcomes. Of course, if all of the conditions for Arrow’s theorem were satisfied, preference orderings would be observable. But they are not satisfied, even when the preference profile is not cyclical (whether or not IIA is satisfied) as soon as any voter votes strategically, and we know from the Gibbard-Satterthwaite theorem that this is always possible if there are at least three alternatives.

Social choice theory is able to entertain the illusion of observability of orderings only by assuming the problem away by assuming that voters never vote strategically. This leads to the assumption that there

---

14 With three alternatives, for example, such an agenda might first pit A and B against each other in a pairwise majority contest, and then the winner against C.
is a one-to-one relationship between a voter's preferences and votes. This assumption, when combined with the use of aggregation rules, gives rise to the idea that different voting rules embody different interpersonal comparisons.

Let me conclude this section with a reflection on how interpersonal comparisons can legitimately be used in studying voting rules. Suppose that a voting theorist studies the relative performance of different voting rules, and makes interpersonal comparisons so as to be able to judge the alternatives normatively. Such comparisons are made in a *methodologically satisfactory way* if they do not favour any voting rules as compared to others. The same interpersonal comparisons can be applied in all voting rules if the same utility profiles are used to study different voting rules. Hence, they can be made in a methodologically satisfactory way.

Consider now a possible way of interpreting this proposal to use interpersonal comparisons of preference intensities in evaluating voting rules. It might be taken to imply that one should make interpersonal comparisons in studying real elections. Suppose, for example, that voting theorists tried to find the *best candidate* in, say, the U.S.A. presidential elections in 2000 by postulating specific cardinal and fully comparable utilities for the millions of voters. It is obvious that such an enterprise would be preposterously misguided because we do not have such information on individual intensities, and we cannot have such information. But then again, we do not have information on preference orderings either. The reason is—as we have seen—that the possibility of strategic voting makes it impossible to obtain reliable information on individual preferences from voting choices: votes are likely to differ from the underlying preferences.

Finding the underlying preferences *from* the votes would require quite some knowledge: since preference intensities have been shown to affect strategic voting, we would need to know what they are, but we already agreed that this is even more difficult than finding out about the preference orderings. But this would not suffice either because we would also need to know the exact distribution of behavioural dispositions because some people refuse to vote strategically even when they have incentives to do so. And as Riker (1982) argued, we have no way of finding out who votes strategically. Finding out about the preference orderings in real elections is thus just as impossible as finding out about interpersonal comparisons. From this perspective,
it is easy to understand why social choice theorists occasionally note that they do not know what the preference profile is in any real election.

These considerations explain why social choice theorists do not use real-world data to evaluate their theories.\(^{15}\) Thus, for example, when Saari (2003) talks about preference profiles as “data”, he refers to profiles that the social choice theorists consider as arguments for an aggregation rule, but these “data” are constructed or postulated by the theorist rather than given by observations. What can be given in observations are the votes.

Whatever is the cogency of observability arguments for ordinal utility in other contexts, the epistemic argument should not have much weight against making interpersonal comparisons of preference intensities in voting theory, because the field is, and will be, dominated by approaches that do not use any kind of real data on actual individual preferences. Social choice theorists thus apply observability criteria selectively: they argue that interpersonal comparisons should not be made on observability grounds, even though they never use any real-world evidence and there cannot be such evidence,\(^{16}\) because there are no reliable observations on preference orderings in the voting context.

In my work on strategic voting (Lehtinen 2007a; 2007b; 2008), the utility profiles are randomly generated, and the interpersonal comparisons are thus also random but the same under every voting rule because the concern the utility profile. Such random interpersonal comparisons are of course methodologically legitimate in the sense indicated above. If one is concerned over the justification for using this or that particular interpersonal comparison, it is surely some consolation that making different interpersonal comparisons of preference intensities with a given ordinal profile does not really affect the results: thus far they have always been robust to different interpersonal comparisons (see Lehtinen 2007b; 2008). Such robustness results indicate that whether one uses this or that particular comparison is unlikely to make a difference to the normative evaluation of voting rules.\(^{17}\)

\(^{15}\) Attempts to measure the amount of strategic voting in some real elections constitute an exception. However, the fact that various authors disagree about this amount is testimony to the underlying observability problem.

\(^{16}\) Tideman and Plassmann (2014), however, use a dataset of actual votes.

\(^{17}\) I do not wish to discourage those interested in voting theory from experimenting with different interpersonal comparisons. It would be an important contribution to show that some voting rule yields satisfactory results only under some peculiar interpersonal comparisons. However, given that interpersonal comparisons do not
5. Debating about Conditions

Let us now have a closer look at the idea of evaluating aggregation rules by the set of conditions that they satisfy. The conditions determine how the outcomes under aggregation rules should depend on individual preferences under any possible configuration of individual preferences, and violations of conditions are demonstrated by way of giving theoretical counterexamples. Given that social choice theorists never (except in the case described below) pass judgment on which alternative or candidate would be the best on the basis of a given utility or preference profile, i.e., independently of the aggregation rules that represent various voting rules, they never have to make interpersonal comparisons. Indeed, if interpersonal comparisons are not to be made, evaluating conditions satisfied by the aggregation rules becomes the only way in which voting can be normatively evaluated.

In contrast, defining the best alternative in terms of a given preference or utility profile requires making interpersonal comparisons. Arrow argues that evaluating the alternatives is tantamount to using a decision-making process: welfare judgement means “an evaluation of the consequences to all individuals based on their evaluations [...] The process of formation of welfare judgements is logically equivalent to a social decision process or constitution” (Arrow 1983 [1967], 68). Since the conditions concern the relationships between individual preferences and the outcomes of aggregation rules, there is no need to pass judgement on which alternative is the best, in any given situation, in terms of the preference profile alone, independently of how the profile affects the aggregation rule. In other words, one never evaluates the alternatives on the basis of their utility (or preference) characteristics in the profile. All the normative work is done, and must be done, by the conditions that specify how the characteristics of any abstract set of individual preferences should translate into social orderings or choices via the aggregation rule.

In social choice theory, the conditions never specify the relationship between the profile of individual preferences and the actual outcomes of voting rules when behavioural assumptions are taken into account. They rather specify the conditions between the preference profile and the aggregation rule under the assumption of sincere voting. The theory

have anything to do with how voters vote or how the voting rules compute the outcomes, I would be rather surprised if someone could establish such results.
proceeds on the assumption that satisfying or failing to satisfy various conditions tells us about the kind of information that is being collected by the different rules. Some rules (e.g., the Borda count) are supposed to obtain information on preference orderings, some others (e.g., agenda voting) on pairwise comparisons, some on approval (approval rule), some on top preferences (e.g., plurality rule), and some on preference intensity (e.g., range voting, the zero-one rule, and perhaps majority judgment).

I will now show that, since strategic voting breaks the connection between preferences and choices, it is no longer valid to infer that a voting rule reflects a certain kind of information if the corresponding aggregation rule satisfies a given condition. The problem is that different voting rules reflect different kinds of information only on the assumption of sincere behaviour. The conditions are imposed on aggregation rules rather than voting rules, but we should be more interested in conditions imposed on the latter because they take behavioural assumptions into account. While different aggregation rules may satisfy different conditions, no voting rule satisfies any condition that makes a difference to the comparative normative evaluation of such rules.

In order to show the relevance of the problem, I will discuss a recent debate between Donald Saari (2000; 2003; 2006) and Mathias Risse (2001; 2004; 2005; 2009a; 2009b) which concerns—at least seemingly—whether the Borda count or the Kemeny rule (or as Risse calls it, the “Condorcet proposal”) is a better voting rule. My main interest does not lie in discussing the relative merits of the two voting rules, but rather in showing with a contemporary normative debate on voting rules how problematic it is to evaluate voting rules on the basis of whether or not the corresponding aggregation rules satisfy various conditions.

Saari’s main line of argument is that one has to be able provide conditions that make a difference between voting rules, and that this can only be done with an appeal to the “data”, the preference profile: “We should try to use the information from the full profile to determine the will of the people” (Saari 2003, 342). If the profile of voters’ preferences has symmetrical elements called Condorcet n-tuples, the voting rule should yield a complete tie with regard to those preferences. Then, removing from the profile voters that are involved in such n-tuples shows the difference-making voters. In other words, any voting rule should provide the same choice or ranking with and without such
voters. Saari presents two new conditions that any voting rule should satisfy. Let ABC denote the ordering A is better than B, and B is better than C. The Neutral Reversal Requirement (NRR) states that pairs of voters with orderings ABC and CBA should be removed from the profile.

A Condorcet triplet (i.e., an n-tuple in the case of three candidates) is constructed as follows. Take the most preferred candidate from the ordering ABC, and place it last so as to get BCA, then do the same to this ordering so as to get CAB. Note that each candidate is at each position exactly once in the triplet {ABC, BCA, CAB}, so that in this sense, no candidate is more preferred than another. The Neutral Condorcet Requirement (NCR) states that such cyclic elements should be removed from the profile. NCR is the condition that makes a difference to the evaluation of Borda vs. Kemeny: The Borda count satisfies it, but the Kemeny rule does not (and on this point Saari and Risse agree).

Consider now the profile that Saari and Merlin (2000) employed to demonstrate that the Kemeny rule does not satisfy the NCR:

<table>
<thead>
<tr>
<th>voter type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of voters</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>A (1)</td>
<td>A</td>
<td>B (1)</td>
<td>B (1)</td>
<td>C</td>
<td>C (1)</td>
<td></td>
</tr>
<tr>
<td>B (.4)</td>
<td>C</td>
<td>A (.1)</td>
<td>C (.5)</td>
<td>B</td>
<td>A (.9)</td>
<td></td>
</tr>
<tr>
<td>C (0)</td>
<td>B</td>
<td>C (0)</td>
<td>A (0)</td>
<td>A</td>
<td>B (0)</td>
<td></td>
</tr>
</tbody>
</table>

The outcomes are determined as follows under the Kemeny rule. If there is a Condorcet winner, the rule selects it. If not, the rule aims to find the best ranking by looking at how many pairwise reversals of preferences would be required to yield a transitive social ranking. In this example, 11 voters prefer A over B, and 8 have the reverse preference. Similarly, there are 9 voters who prefer A to C, and 10 with the reverse preferences; and 14 voters who prefer B over C, and 5 voters with reverse preferences. These pairwise preferences constitute a cycle {AB, BC, CA}, and the social ordering is ABC because it requires that only one type 6 voter reverses her ordering of A and C.

There are two Condorcet triplets. {ABC, BCA, CAB} constitute one cycle and {ACB, CBA, BAC} another. The number of voters with orderings ABC, BCA and CAB are 6, 5, and 5, respectively. Since the smallest of these numbers is 5, we remove 5 voters from these three voter types so that the remaining profile is the following.
Table 7: Applying the NCR

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>A (1)</td>
<td>A</td>
<td>B (1)</td>
<td>B</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>B (.4)</td>
<td>C</td>
<td>A (.1)</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>C (0)</td>
<td>B</td>
<td>C (0)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
</tbody>
</table>

BAC is thus the social ordering after removing the cyclic elements, but this means that the NCR is violated. However, consider now strategic voting. If type 3 voters realise that sincere voting is likely to yield ABC, they may vote strategically by reporting BCA rather than BAC so as to lower A’s position. If they do, the reported rankings are:

Table 8: Strategic orderings

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
</tbody>
</table>

Then the pairwise comparisons are:

A over B: 11 vs. 8
A over C: 6 vs. 13
B over C: 14 vs. 5

The social ranking is now BCA because this ordering only requires reversing two voters’ AB preference, the social choice would be B, and type 3 voters would be satisfied. In other words, the Kemeny rule does not violate the NCR with this profile if voters of type 3 act strategically in this way.

Consider now the Borda count. The Borda scores are computed as follows. With n candidates, the most preferred candidate obtains n-1 points, the second-best n-2,..., and the least preferred 0 points. The Borda scores are 20, 22, and 15 for A, B, and C, respectively, so that B is selected. However, if, say, three type 6 voters vote strategically by reporting ACB rather than CAB, the Borda scores are 25, 22, and 10 for A, B, and C, respectively, and A is selected instead of B. But now the
Borda count no longer satisfies the NCR because removing the voters within the Condorcet triplet \{ABC, BCA, CAB\} yields B as the winner. Note, however, that type 3 voters could thwart such a strategy because they could report BCA instead of BAC:

| Table 9: Strategic orderings under the Borda count |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1 2 3 4 5 6     |
| 6 3 0 8 0 2     |
| A A B B C C     |
| B C A C B A     |
| C B C A A B     |

If they do, the Borda scores are again 20, 22, and 15, for A, B, and C, respectively. Suppose, however, that they do not realise that they should do this because all type 6 voters seem to speak in favour of C.

To summarize, we have seen that the Kemeny aggregation rule violates the NCR, that the Kemeny voting rule does not violate the NCR, that the Borda aggregation rule does not violate the NCR, and that the Borda voting rule violates the NCR, and all these results were derived from a single profile. What difference does it make, then, that the Borda aggregation rule satisfies the NCR, but the Kemeny aggregation rule does not? This is not an argument for Kemeny or against Borda because, as far as I understand, Risse’s main argument is that the Kemeny rule finds the “best ranking”. If ABC is the “best ranking”, the Kemeny voting rule does not find it in our example with strategic voting.

Breaking the connection between preferences and choices through strategic behaviour means that satisfying a condition does not necessarily make a voting rule normatively more attractive than another. If the conditions that are currently imposed on aggregation rules were to be formulated in terms of the relationship between the preference profile and the actual voting outcomes, all of the conditions that make a difference between different voting rules could be violated in any voting rule.

There are conditions like anonymity, neutrality, and universal domain, which cannot be violated through strategic voting, and which most voting rules satisfy. It is clear that my claim does not concern such conditions. Tideman (2006, 151) provides a classification of five different kinds of condition: domain, consistency, responsiveness, stability, and qualitative attractiveness. My claim concerns conditions in
the consistency (e.g., Condorcet, majority, transitivity), responsiveness (e.g., positive responsiveness), and stability (e.g., monotonicity and various independence conditions, perhaps strategy-proofness) categories.

Social choice theory does provide some useful information about the functioning of voting rules because, for example, some rules may only yield satisfactory results with single-peaked preferences. Violations of neutrality and anonymity may well be important for some questions. However, I claim that social choice theorists’ disagreements over the most common voting rules can be traced back to different views concerning the relative importance of various difference-making conditions. My point is that even if one could agree on their relative importance, whether a given condition is satisfied by some aggregation is irrelevant because the corresponding voting rule will always violate it.

Although I have not provided an exhaustive list of examples in which all the various conditions are violated, such examples can easily be generated for any given rule and condition. Violating a given condition under given voting rule with a specific preference profile always requires that at least two alternatives are at least somewhat popular. If a profile of ordinal preferences is proposed and some alternative, say A, is shown to be selected under sincere behaviour, one can always find a way in which another alternative B becomes selected by introducing high (intrapersonal) preference intensities for B and low intensities for A. The reason for this is that in all voting rules, alternatives for which intensities are high on average (i.e., for which the sum of utility is large) are more likely to obtain strategic votes than other alternatives (Lehtinen 2007a; 2007b; 2008).

It is impossible to provide a general proof of my claim because the conditions concerning utilities and beliefs under which voters act strategically are different under each rule. Yet I am not willing to withdraw from the strong formulation of the claim that involves all voting rules and all difference-making conditions because in every rule that has been studied thus far, the utilitarian winner obtains a lot of strategic votes. Furthermore, given that strategic voting always depends on preference intensities and on voters’ incomplete information, it is to be expected that similar results hold under all voting rules.

---

18 Suppose that there are 20 genuinely different voting rules and 20 genuinely different difference-making conditions. It would then take 400 examples of the kind described in this paper to prove that all difference-making conditions can be violated.
These results mean that, as long as the conditions are formulated in terms of ordinal preferences, one can always generate a violation of any condition (which is formulated between real preferences and outcomes of voting rules) by tweaking voters’ intensities or behavioural dispositions. I thus challenge social choice theorists to come up with just one difference-making condition and a voting rule in which that condition is not violated.19

One might argue that evaluating voting rules on the basis of conditions that the corresponding aggregation rules satisfy makes sense despite my criticism because there are cases in which a condition is violated only under strategic behaviour, but also under sincere behaviour under another voting rule. Thus, a given condition may be more frequently violated under one rule than under another. However, even this argument is questionable because a voting rule may violate a condition under sincere behaviour with a given preference profile, but satisfy it under strategic behaviour with exactly the same preference profile. We have already seen an example of this with the Kemeny rule. For another example consider a violation of monotonicity from Tideman (2006, 191). 11 voters have the following preferences among three alternatives A, B, and C. I have supplemented the preference intensities in parentheses so as to provide a plausible account of strategic voting.

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>4</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(1)</td>
<td>B</td>
<td>(1)</td>
<td>C</td>
</tr>
<tr>
<td>B</td>
<td>(.9)</td>
<td>C</td>
<td>(.9)</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>(0)</td>
<td>A</td>
<td>(0)</td>
<td>B</td>
</tr>
</tbody>
</table>

Under the runoff rule, two alternatives obtaining the most votes in a first round of voting meet in a second-round pairwise contest (unless

---

19 It may be tricky for me to provide examples for some rules under which defining the strategic voting strategy is particularly complex. To the best of my knowledge, there are no models of strategic voting under the Single transferable vote (STV) and the Majority judgment. One way to deal with STV is to use a “plurality heuristic”: assume that voters behave as if the voting rule were the Plurality rule, and use McKelvey and Ordeshook’s (1972) pivot probability model to define expected gains for each alternative. The strategic rankings are then given by the ranking of expected gains. I have obtained some simulation results (under STV) which indicate that such behaviour is rational in the sense that, on average, voters that give such rankings are better off than they would be under sincere behaviour. At present, I do not know whether it makes sense to use the plurality heuristic in the majority judgment.
one candidate has a majority in the first round). This aggregation rule violates monotonicity:\textsuperscript{20} A is eliminated in the first round of voting, and B wins the second round against C. However, if the voter in the last column puts B ahead of C, A and C tie for fewest votes, and if the tie is resolved in A’s favour, it also wins B in the second round. Note carefully, however, that if this voter were to realise the consequences of her change of mind, she would see that changing her vote is against her own interests: instead of the tolerable B, the social choice will be the abhorred A, and she should vote strategically for C. If she does, B continues to win. In other words, strategic voting implies that monotonicity is no longer violated. The runoff aggregation rule thus violates monotonicity with this profile, but the runoff voting rule does not. Note that B beats C rather clearly 7-4 in the second round, and that the race between A and C is close in the first round. This voter’s strategic choice is thus quite sensible.

Consider now the Borda aggregation rule which satisfies monotonicity. The Borda scores are 9 for A, and 12 for B and C. B and C thus tie for victory. If one of the four voters in the second column puts C ahead of B by reporting CBA, C becomes the sole winner and monotonicity is satisfied. What happens under strategic behaviour? If the three voters in the leftmost column vote strategically by reporting BAC, they avoid the abhorred choice of C. The Borda scores would then be 6, 14, and 12, for A, B, and C, respectively. B would thus become the sole winner and the Borda voting rule violates monotonicity. Again, the irony lies not in the fact that an alternative that has become more popular is no longer chosen, but rather in the fact that this happens in a voting rule that corresponds to an aggregation rule that satisfies \textit{monotonicity}. It is true that the voters could vote strategically in this way even before the BCA voter changes her preference ordering. Note, however, that they have \textit{systematically} stronger incentives to do so after the change: rather than a tie between B and C, the social choice would now be C under sincere behaviour. What difference does it make that the Borda social choice rule satisfies monotonicity and the runoff does not?

\textsuperscript{20} A ranked voting system is \textit{monotonic} if it is neither possible to prevent the election of a candidate by ranking them higher on some of the ballots, nor possible to elect an otherwise unelected candidate by ranking them lower on some of the ballots (while nothing else is altered on any ballot). In single winner elections that is to say no winner is harmed by up-ranking and no loser can win by down-ranking.
A voting rule $V_1$ cannot be better than $V_2$ on the grounds that the corresponding aggregation rule $A_1$, unlike $A_2$, satisfies a normatively compelling condition $C$ if voting rule $V_1$ may violate $C$ and voting rule $V_2$ satisfy $C$ with exactly the same profile. As a result, social choice theory is unable to provide useful normative information on the relative merits of various voting rules.

Given the ease with which such examples can be generated, whether or not an aggregation rule satisfies a given condition does not really seem to matter. We do not seem to have any reason to assume that if a condition is violated under aggregation rule $A_2$ and satisfied under aggregation rule $A_1$, it is more frequently violated under the corresponding voting rule $V_2$ than under $V_1$. Comparative normative advice which is derived from social choice theory is always uncertain. All we can say is that no real voting rule ever satisfies the intuitive demands behind any condition that makes a difference. Social choice theory is able to entertain the illusion of being able to provide normative advice only by assuming away what prevents it, viz., by assuming that voters never vote strategically. We do not know which rule is more likely to violate this condition in real voting rules because nobody has studied the issue with a plausible model of strategic voting.\footnote{It can be done, however. I have studied path-dependence, see Lehtinen 2015. Felsenthal provides such results under several rules, but he assumes that strategic voters have complete information, see Felsenthal 1990.} I conclude that it is deeply problematic to evaluate voting rules in terms of the sets of normative conditions that they satisfy under the assumption of sincere behaviour because the inferences from social choice rules to real voting rules are highly unreliable due to strategic voting.

Different voting rules may be normatively compared as soon as we are willing to pass judgements on the alternatives merely on the basis of the utility profile. The best voting rules are then found simply by seeing which ones select the best alternative, i.e., the utilitarian winner, most frequently. This kind of evaluation, however, requires rejecting two important principles of social choice theory. First, in order to evaluate the alternatives on the basis of the utility profile, one must be willing to make interpersonal comparisons of utility that concern the choice alternatives. Secondly, one must distinguish between preference and choice so as to be able to define different voter behaviours in different rules while retaining the same utility profile. The normative standard
must be behaviour-independent because otherwise outcomes from either different voting rules or different behavioural assumptions cannot be compared. Comparability also requires that one must keep the utility profile fixed across voting rules.

In social choice theory, the standard for evaluating whether or not a given voting rule violates a given condition is whether one can find an example of a preference profile that violates it. There is no concern over how alarmingly unlikely it is that a condition is violated. An obvious remedy for this is to study many utility profiles. Studying thousands or millions of profiles across voting rules is possible, however, only with a computer simulation, and this will require rejecting another principle of social choice theory: that only analytical results are considered acceptable.

6. STRATEGIC VOTING IS BENEFICIAL

We have seen that taking strategic voting into account destroys the normative force of social choice theory in making comparative assessments of voting rules. Consider now the second main topic of social choice theory, namely, evaluating the general performance of voting rules in collecting information on individual preferences. Arrow's theorem and the aggregation functions have been interpreted as being concerned with social decision mechanisms such as voting rules, and with social welfare judgements (see Buchanan 1954; Sen 1977; 1995). Accordingly, there are two main interpretations of the concept of an aggregation rule and of Arrow's theorem. The first is that there is no optimal voting rule because cyclic preferences make path-dependent social choices, strategic voting, and agenda manipulation possible (e.g., Bordes and Tideman 1991). The theorem has thus been interpreted to mean that there is no ideal democratic system of voting. The second interpretation is that there is no such thing as a well-defined "social optimum" (e.g., Riker 1982, 137; Ordeshook 1986, 57). Sen's 1970 proof of an impossibility theorem with cardinal utilities is commonly taken to

22 There may be quantitative differences in how often various rules violate the various conditions, and some contributions have tried to compute violation probabilities from a variety of different profiles. For example of this kind of work, see Mbih and Moyouwou 2008. See also Nurmi 2012.

23 Admittedly, the impartial culture assumption, which is used in many models that study strategic voting, is very likely to generate more strategic voting opportunities than there is in reality. However, since some studies that measure the extent of strategic voting indicate somewhat extensive strategic voting (e.g., Blais, et al. 2001), it would be odd to deny its importance on the basis of the data that we have.
imply that it is impossible to define a social optimum without making interpersonal comparisons.

Riker (1982) argued that the possibility of strategic voting renders voting results meaningless. More generally, Arrow's theorem and the Gibbard-Satterthwaite theorem are often interpreted as implying that there is no democratic voting rule that satisfies a set of intuitively acceptable conditions, and that all methods of aggregating individual preferences are unsatisfactory. The justification of democracy thus rests on a shaky conceptual foundation if this is the right interpretation of the theorems.

If there is anything that social choice theorists agree on, it is that strategic voting should be minimized if possible. Furthermore, it appears that social choice theorists have also managed to convince other political scientists and philosophers (e.g., Miller 1992; Dryzek and List 2002), and even those who are not sympathetic to social choice theory (Christiano 1995), that only distortion and chaos can come from strategic voting. The Gibbard-Satterthwaite theorem states that any single-valued strategy-proof voting procedure that considers at least three alternatives is dictatorial. A voting procedure satisfies strategy-proofness if no individual has an incentive to vote strategically.

While there is disagreement concerning which particular conditions are the most important, the general welfarist principle that the best voting rules should generate outcomes that reflect individual preferences is universally accepted. Yet, the arguments that have been presented in favour of strategy-proofness have been non-welfarist. The arguments are non-welfarist because evaluating the welfare consequences of strategic voting explicitly requires making interpersonal comparisons. Any such evaluation must postulate two different behavioural assumptions, sincere and strategic, and a behaviour-independent utility-based normative standard with respect to which the outcomes may be compared. It is impossible to evaluate the welfare consequences of strategic voting without making interpersonal comparisons of voter utilities because the normative standard must be independent of voters’ behaviour and rather based on the utility profile.

Cohen (1986) points out that Bentham distinguished between the utility principle and the majority opinions, believing that “general consent” provides the surest visible sign and immediate evidence of the

---

24 For these arguments, see Kelly 1988, 103; and for an extensive critique, see van Hees and Dowding 2008.
sum of utility. I also take the social optimum to be given by the utilitarian winner and adopt an epistemic populist notion of voting: voting provides fallible evidence of the best candidate. Note, however, that Coleman and Ferejohn (1986) as well as Cohen argue that a unique general does not always exist. I have assumed that the utilitarian winner always defines the best outcome, and that the interesting question is whether, how often, and under which conditions the various voting rules find it. I promote adopting this stronger interpretation of epistemic populism in evaluating the performance of voting rules.

Studying the welfare consequences of strategic behaviour requires a kind of counterfactual comparison. One must deduce sincere and strategic behaviour from a single utility profile, and compare the outcomes against the normative standard. Such an exercise cannot be conducted in a traditional social-choice framework in which individual utilities or preferences are aggregated because aggregation does not allow the aforementioned disentangling of utilities from choices. In other words, studying the welfare consequences of strategic voting in a welfarist manner requires incorporating epistemic populism into the very structure of the model with which the evaluation is made.

As we have seen, strategic voting increases the frequency with which the utilitarian winner is chosen in all major voting rules (Lehtinen 2007a; 2007b; 2008). Trying to minimize strategic voting might lead to worse outcomes in terms of individual utilities. The most unanimously cherished condition among social choice theorists is then not normatively acceptable from the utilitarian point of view. The results of social choice theory with respect to strategy-proofness provide a fundamentally misleading picture of the performance of various voting rules in finding satisfactory alternatives in terms of individual utilities, both comparatively and absolutely.

Note also that the methods of social choice theory could not have produced the result that strategic voting is beneficial because it is an emergent outcome: it makes sense to impose conditions on the relationships between inputs and outputs in an aggregation only if we know that there are no emergent properties (see, e.g., Wimsatt 1997), but the very idea of aggregation is misguided in the context of voting because an invisible hand affects the outcomes (see Lehtinen 2009).

Consider now the consequences of the fact that strategic voting is beneficial from the point of view of Arrow’s theorem. Given that all social choice results assume sincere behaviour, there cannot be an
**acceptable set of conditions** which is satisfied by any real-world voting rule if strategic voting is beneficial: Strategic voting implies that no difference-making condition can be satisfied by voting rules, but at the same time, the results of voting would be worse in utilitarian terms in the absence of strategic voting. Arrow’s set of conditions, or any set of conditions that has been proposed in social choice theory cannot be normatively acceptable because it always includes the assumption that voters act sincerely.\(^\text{25}\) This means that while Arrow’s theorem and the Gibbard-Satterthwaite are logically impeccable, they fail to have the devastating consequences for democracy that have sometimes been attributed to them.

Donald Saari’s (1995; 1998; 2010) arguments against interpreting Arrow’s theorem too strongly are similar to mine in spirit but he targets a different condition. He argues that IIA is not normatively acceptable because it overrides transitivity and thus individual rationality. Satisfying IIA is problematic because it means that an aggregation rule must ignore information on the transitivity of preferences even though transitivity has been explicitly required, and thus that aggregation rules that satisfy IIA fail to recognise rational voters. Saari’s conclusion is powerful if one accepts the standard aggregation rule framework. However, I do not think this argument has much force when considering voting rules. First, fortunately, rational voters can resort to strategic voting, and since rational voting always requires evaluating at least three alternatives, even voting rules that correspond to IIA-satisfying aggregation rules may indirectly take at least some transitivity information into account. Secondly, strategic voting under voting rules that correspond to aggregation rules not satisfying IIA may often mean that the orderings reported do not correspond to the actual preference orderings. Thus, even if a voting rule explicitly collects information in terms of orderings, there is no guarantee that it takes the correct transitivity information into account.

Note that if there is any condition that will be violated as a consequence of strategic voting, it is individual rationality (see Lehtinen 2011 for details). If individual i’s choice from a set of alternatives S is C\(_i\)(S), and xR\(_i\)y denotes her preferences, individual rationality can be expressed as follows: C\(_i\)(S)={x|x∈S: ∀y∈S: xR\(_i\)y}. It is this condition that is always violated, and if one agrees with me that strategic voting is very

---

\(^{25}\) This is usually expressed in terms of *individual rationality*. 

---
likely to be rather commonly beneficial, it is this condition that is not normatively acceptable.

Social choice theory has evolved into *implementation theory.*\(^{26}\) This theory studies which normative conditions various institutions satisfy in various game-theoretical equilibria. The theory thus relaxes the assumption of sincere behaviour while retaining the endeavour to study normative issues by way of examining whether institutions (such as voting rules) satisfy various normative conditions.

Insofar as the theory is used to study voting, the aggregation rules\(^ {27}\) are taken to characterize the desirable properties of voting institutions. Therefore, just like in social choice theory, interpersonal comparisons need not be made because the very same conditions are studied and the theory never passes any normative judgments on a particular profile. The theory rather defines the desirable outcomes in terms of aggregation rules which satisfy various conditions. The basic idea is to study whether one can find a *game form* which always yields the same outcomes as those that sincere voting would generate under the aggregation rule in question. A game form, which is also sometimes called a ‘mechanism’, is a game without the players’ preferences: it specifies the strategies available for the players and the consequences of various combinations of strategies. An aggregation rule is said to be implementable in \(\beta\)-equilibrium if, for any profile of preferences, there is a game form in which the strategic behaviour of individuals leads to the same outcomes as those that would ensue from sincere voting in the aggregation rule. Here \(\beta\) refers to the various possible solution concepts. For example, if the Borda aggregation rule were implementable with dominant strategies, then this would mean that the conditions satisfied by this rule would always be satisfied in a game form in which the players have dominant strategies.

The most celebrated result in this field is the so-called *revelation principle* (Gibbard 1973; Myerson 1979) which posits something like this: If an institution implements an aggregation rule, then there is a *direct mechanism* which also implements that rule. A direct mechanism is a game form in which truth-telling is equilibrium behaviour. To put it simply, and relating the point to voting institutions, the revelation principle posits that if there is an aggregation rule that is implemented

\(^{26}\) For introductions to this theory, see Jackson 2001; or Austen-Smith and Banks 2005, chapter 3.

\(^{27}\) These are often called ‘social choice rules’ in implementation theory.
by a game form (voting institution), then that aggregation rule can be implemented in sincere strategies. In other words, if some set of normative conditions can be satisfied in equilibrium, then those normative conditions can be satisfied in a game form in which the players’ truth-telling is equilibrium behaviour.

The revelation principle is similar in spirit to what I have said about violating conditions in that it indicates that sincere voting at least can be compatible with satisfying various conditions, and I have argued that strategic voting implies that all other relevant conditions will be violated. It is this incompatibility of strategic voting with any condition together with the beneficial nature of strategic voting that generates the impossibility of a set of normatively acceptable conditions. Given that strategic voting implies that all other difference-making conditions will be violated, it is perhaps not so surprising that none of the aggregation rules that are commonly used in political elections are implementable in Nash equilibrium (see, e.g., Jackson 2001, 672). This is the reason why implementation theory is just as incapable of providing comparative normative advice on voting rules as social choice theory. This impasse in implementation theory may explain in part why scholars working on voting and welfare may have chosen to bite the bullet and continue studying normative issues related to voting with the traditional tools of social choice theory. Ultimately, this impasse derives from the same problem that afflicts social choice theory: strategic voting is incompatible with every other normatively desirable condition that aggregation functions might satisfy.

Note that the result that strategic voting increases the frequency with which the utilitarian winner is selected is intuitively in conflict with the revelation principle because it means that strategic voting may be necessary for obtaining normatively desirable results. Implementation theory would study a similar question by looking at whether the zero-one rule is implementable under some solution concept. It is clear that it is not implementable under any solution concept because full implementation would require that the utilitarian winner is always selected. This reminds us once again that satisfying just about any condition is too much to ask, and that obtaining genuinely useful results for a comparative normative evaluation of voting rules may require computer simulations.

One might also argue that any given set of conditions that precludes taking intensity of preference into account is not normatively
acceptable. I agree with social choice theorists that there is no optimal voting rule because there is no rule that is guaranteed to find the utilitarian winner. The correct conclusion from this is that trying to propose a set of conditions that would take preference intensities into account is also doomed to failure: no voting rule will satisfy a set of conditions that take intensities into account because strategic voting implies bullet voting under the zero-one rule or range voting, and this means that it is impossible to obtain intensity information reliably from those rules as well. Intensities are thus something that cannot be properly treated by imposing conditions on aggregation rules or on voting rules. However, they are out there, and the only way in which we can give them their proper normative importance is by making interpersonal comparisons.

7. What kind of interpersonal comparisons should one make?

Let us finally consider what the interpersonal comparisons should be like in voting theory. In order to do so, let us go back to the dispute between Saari and Risse. Table 6 is reproduced here for ease of reference.

<table>
<thead>
<tr>
<th>voter type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of voters</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>A (1)</td>
<td>A B (1)</td>
<td>B (1)</td>
<td>C</td>
<td>C (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B (.4)</td>
<td>C A (.1)</td>
<td>C (.5)</td>
<td>B</td>
<td>A (.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C (0)</td>
<td>B C (0)</td>
<td>A (0)</td>
<td>A</td>
<td>B (0)</td>
<td></td>
</tr>
</tbody>
</table>

Consider first whether violations of NCR are desirable in the example we considered in section 5. The sums of utilities are 6+(3*.1)+(5*.9)=10.8 for A, (6*.4)+3+5=10.4 for B, and (5*.5)+5=7.5 for C in the initial situation, and 1.1 for A, 3.4 for B, and 0 for C in the final one. In other words, according to the utilitarian criterion, A should be selected in the initial situation and B in the final one. Violating the NCR through strategic voting in the Borda voting rule is desirable because the utilitarian winner A becomes selected. The utilitarian criterion thus conflicts with the NCR, and for obvious reasons: the NCR neglects the preference intensity information by treating all identical orderings the same. Saari argues for the NCR on the grounds that profile settings where no candidate has an advantage over another should yield
a complete tie because it “reflects the principle that each voter’s preferences are treated equally” (Saari 2006, 112). This is an interpersonal comparison because it concerns the preference profile. This means that Saari’s conditions are based on interpersonal comparisons of ordinal utilities: if the voter with ordering ABC has a much greater stake in the issue than the other two voters with orderings BCA and CAB, it would seem inappropriate to declare that the social choice should be a complete tie: there would be an argument for selecting A (see, e.g., Brighouse and Fleurbaey 2010).

In sum, if one agrees that preference intensities matter, violating NCR seems to be perfectly justified, and the fact that the Borda voting rule violates NCR is not meant to be an argument against the rule. I am somewhat diffident to present these remarks on the NCR as a criticism of Saari, because he is clearly aware of the fact that the whole debate becomes different if preference intensities are taken into account (2003, 349). Furthermore, as far as I know, he has been silent on preference intensities—except “intensities for rankings” (see, e.g., Saari 1995)—and their interpersonal comparisons.

Saari’s admonition to treat each voter’s preferences equally may be compared to Hammond’s claim for the zero-one rule. He argues that “it seems almost perverse not to make the weights equal” (Hammond 1987, 197). Indeed, that is what we have done thus far by assuming that each voter’s maximum utility is unity and the minimum zero. However, James Griffin argues that the zero-one rule is, in fact, empirically false: “It is not the case that we all reach the same peaks and valleys” (Griffin 1986, 120). In other words, if zero-one normalisation is considered an empirical assumption about people’s utilities, it is false. If Griffin’s empirical claim is correct (remember, we do not have any scientific means for determining whether it is), insofar as each person still has only one vote, the conclusion is that voting rules cannot take empirical interpersonal differences in utilities into account (see Mackie 2003, 144-145). This example could be analysed by assuming that voters’ utility numbers are drawn randomly from the (0, 1) interval. If interpersonal comparisons are introduced in this way, the conclusions of the analysis are likely to be the same as with the uniform zero-one assumption. How about assuming that type 3 voters have double the stakes compared to types 4 and 6, and that type 1 voters have half the stakes? The utilities would then be as follows:
Table 11: Modified example

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A (.5)</td>
<td>A</td>
<td>B (.2)</td>
<td>B (1)</td>
<td>C</td>
<td>C (1)</td>
<td></td>
</tr>
<tr>
<td>B (.2)</td>
<td>C</td>
<td>A (.2)</td>
<td>C (.5)</td>
<td>B</td>
<td>A (.9)</td>
<td></td>
</tr>
<tr>
<td>C (0)</td>
<td>B</td>
<td>C (0)</td>
<td>A (0)</td>
<td>A</td>
<td>B (0)</td>
<td></td>
</tr>
</tbody>
</table>

If we knew that individual utilities are like this, I presume that we would not want to apply the NCR. It would seem sensible to give type 1 voters one half the weight of the other types in the Condorcet triplet \{ABC, BCA, CAB\}. Then, it would be as if there were three type 1 voters with a weight of one. Similarly, it would be as if there were six type 3 voters. The profile that would follow from deleting the voter types in the Condorcet triplet would be as follows:

Table 12: Deleting Condorcet triplets

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>A (1)</td>
<td>A</td>
<td>B (1)</td>
<td>B (1)</td>
<td>C</td>
<td>C (1)</td>
<td></td>
</tr>
<tr>
<td>B (.4)</td>
<td>C</td>
<td>A (.1)</td>
<td>C (.5)</td>
<td>B</td>
<td>A (.9)</td>
<td></td>
</tr>
<tr>
<td>C (0)</td>
<td>B</td>
<td>C (0)</td>
<td>A (0)</td>
<td>A</td>
<td>B (0)</td>
<td></td>
</tr>
</tbody>
</table>

B would now clearly be the best candidate. It is clear that we could not come up with such exact knowledge of interpersonal differences in utilities. My point is just that deviating from the assumption of uniform interpersonal utility scales would be problematic for applying the NCR, but a utilitarian evaluation does not depend on such an assumption. Nevertheless, I do not intend to criticise the NCR for making such a uniformity assumption. No voting rule can take interpersonal differences in utility into account, and normative evaluations of voting rules should employ the same, or at least roughly the same scales.

The one-man-one-vote principle posits that each voter should have only one vote. It is possible to interpret this principle merely in terms of equality of opportunity (e.g., McGann 2006), i.e., in such a way that it does not imply anything about the definition of a social optimum. However, if voting outcomes are interpreted as reflecting individual utilities, they need to be normatively evaluated, and thus the need to
justify the principle in terms of claims about the utility profile arises. In other words, if we are primarily interested in voting results from the welfarist point of view, we also need to justify the one-man-one-vote principle by welfarist means.

If the principle is justified by appealing to interpersonal comparisons, the weight of each individual in determining the social optimum ought to be the same. The justification of the one-man-one-vote principle derives here from the claim that, *a priori*, each individual ought to have equal weight in determining the will of the people. It follows that each individual ought to have the same opportunity to affect the outcome of a voting process. Another way to look at the issue is to note that since the one-man-one-vote principle is violated only if we know that some voters have a legitimate claim to more than equal influence on the voting outcome, when there are no such reasons to violate the principle, we should also assume that each voter's utility is measured with the same scale. We thus impose restrictions on the utility profile by requiring that every voter has the same scale. Even though the individual utilities are usually considered empirically given facts under utilitarianism, here they are also determined by normative considerations. This viewpoint is not entirely new to utilitarianism (see Harsanyi 1975; and Nunan 1981).

Interpersonal comparisons are made in a normatively satisfactory way if the comparisons correspond to acceptable ethical judgements concerning how the utilities of different individuals should be weighed. In particular, if a model is based on interpersonal comparisons, they should be impartial in the right way: particular persons should not be favoured unless there are convincing arguments for doing so. Normatively satisfactory interpersonal comparisons require that the weight of each individual voter in the welfare function that evaluates the outcomes must be equal or at least roughly equal. It follows from the normative justification of the one-man-one-vote principle that the differences should not be too large. No voter’s utility should count more than twice as much as that of any other voter. If it were otherwise, one would wonder why each voter is given one vote. Interpersonal comparisons are thus normatively satisfactory if they take some differences in utility into account, but at the same time, restricting the size of the differences can be justified by a welfarist reading of the one-man-one-vote principle.
8. Conclusions

I have argued that social choice theory does not distinguish between choices and utilities (or preferences), and that this leads them to the mistaken idea that some but not all voting rules incorporate interpersonal comparisons. Strategic voting means, however, that one cannot assume that the ballots cast correspond to voter utilities. Interpersonal comparisons are different in different voting rules only under the questionable assumption that voters are sincere. Once we distinguish between votes and utilities, we can see that different voting rules depend on different kinds of information if voters act sincerely, but once strategic behaviour is taken into account, they all depend on at least some intensity information. Strategic voting also implies that social choice theorists are merely fooling themselves if they claim to have reliable information on ordinal preferences. Strategic voting demotes the epistemic argument against interpersonal comparisons, because it shows that social choice theorists do not and cannot have reliable knowledge of preference orderings either.

If we are not allowed to make interpersonal comparisons then imposing conditions on aggregation rules is the only way in which we can evaluate voting and voting rules. However, strategic voting also has another devastating consequence for social choice theory: it renders the theory unable to provide comparative normative advice on which voting rules are the best. While aggregation rules differ with respect to the conditions they satisfy, no difference-making condition is satisfied by any voting rule. One may provide normative advice if one is willing to make normative judgements about the alternatives in terms of the preference profile alone, rather than in terms of how the profile may affect some function that is supposed to describe a voting rule. Doing so in a welfarist manner, however, requires making interpersonal comparisons. This should not be such a big problem, however, because interpersonal comparisons can be made in a perfectly satisfactory way.

References


Tideman, Nicolaus T., and Florenz Plassmann. 2014. Which voting rule is most likely to choose the "best" candidate? *Public Choice*, 158 (3): 331-357.

**Aki Lehtinen** is a professor (temporary) of Social and Moral Philosophy at the University of Helsinki, and a university researcher at the Finnish Centre of Excellence in the Philosophy of the Social Sciences (TINT). He has written on modelling and robustness, game theory, and other miscellaneous topics in the philosophy of economics. He has also constructed computer simulation models of strategic voting and published the results in economics and political science journals.
Contact e-mail: <aki.lehtinen@helsinki.fi>
Website: <http://www.mv.helsinki.fi/home/alehtine/akilehtinen.htm>

JEFFREY R. DI LEO

*University of Houston-Victoria*

Neoliberalism has been broadly accepted as a fairly recent economic and political project. For example, David Harvey, distinguished professor of anthropology at the Graduate Center of the City University of New York, in his widely cited book, *A brief history of neoliberalism* (2005) looks upon the years 1978-1980 to begin his social and economic history of neoliberalism. For Harvey, key figures from this period, including Deng Xiaoping of China, Margaret Thatcher of Britain, and Ronald Reagan of the United States set the stage for “a revolutionary turning-point in the world’s social and economic history” (p. 1). Others mark the turning point for the project of neoliberalism to the work of Milton Friedman and the emergence of the Chicago school of economics in the 1960s. Still others trace it back to the work of Friedrich Hayek and Lionel Robbins and the London School of Economics during the 1930s. It is within this context of “standard” neoliberal economic and political history that the story told by William Davies stands out.

In a bold and intriguing move, Davies places the foundations of neoliberalism in the late eighteenth-century social and political philosophy of Jeremy Bentham. Davies reminds us that the English philosopher’s hedonism had strong connections with business, government, and the market—a point often overlooked in the rush to dismiss Bentham’s hedonistic utilitarian ethics as merely a philosophically weak precursor to the more philosophically mature eudaimonistic utilitarianism of his student John Stuart Mill. “The business of government”, wrote Bentham in *The principles of morals and legislation* (1789), “is to promote the happiness of society, by punishing and rewarding” (cited by Davies, p. 19). “The free market, of which Bentham was an unabashed supporter, would largely take care of the reward part of this ‘business’”—comments Davies—“the state would take responsibility for the former part” (p. 19).
Thus begins Davies creative and convincing journey from the hedonic calculus and surveillance state of Bentham to the contemporary “happiness industry” and neoliberal state. But Davies’s project is much more than merely parsing out some of the originary moments of neoliberal thought in the happiness science of Bentham, it is also to looking beyond the current formation of neoliberalism to its next position, the post-neoliberal era.

Referring to Hayek’s *The road to serfdom* (1944), Davies notes that “[o]ne of the foundational arguments in favour of the market was that it served as a vast sensory device, capturing millions of individual desires, opinions and values, and converted these into prices” (p. 10). However, for Davies, we may be “on the cusp of a new post-neoliberal era in which the market is no longer the primary tool for this capture of mass sentiment” (pp. 10-11). “Once happiness monitoring tools flood our everyday lives”, writes Davies, “other ways of qualifying feelings in real time are emerging that can extend even further into our lives than markets” (p. 11).

It is here, however, that Davies arguments concerning business, government, and the market go well beyond the standard critiques of neoliberalism and the surveillance state—and extend into the fields of psychology, neuroscience, and health care. To be sure, Davies is very hard on all three. For him, governments and corporations have become obsessed with measuring how people feel and then cashing in on it. The measurement and commercialization of our feelings and emotions through “smart technology”, for example, is clearly not something that Davies thinks really improves our “well-being”. Rather, it is only part of a larger effort to cash in on our emotions and place them under continuous surveillance. “Any critique of ubiquitous surveillance”, argues Davies, “must now include a critique of the maximization of well-being, even at the risk of being less healthy, happy, and wealthy” (p. 11). But is this what people really want? Namely, to be less healthy, happy, and wealthy?

While the eight chapters of the book never really answer this last question, they do a remarkable job of taking us on a two hundred plus year journey from the birth of a science of happiness in the eighteenth century philosophy of Bentham through the various psychologies and sciences of happiness in the nineteenth century to its commercialization in the late twentieth and early twenty-first centuries. The interweaving
of philosophy, psychology, and economics in the book is elegant and fast-paced, and is replete with many fascinating historical facts.

Chapter one, “Knowing How You Feel”, begins with a survey of Bentham’s notion of happiness as a “physical occurrence within the human body” (p. 20) and his thoughts on how it might be measured. Davies offers that Bentham provides two responses to the question, “How does utility manifest itself in such a way that it can be grasped by measurement?” (p. 24). The first is “human pulse rate” (p. 25), a response that Bentham “wasn’t particularly taken with himself” (p. 25); the second was, “that money might be used” (p. 25). Notes Davies, Bentham “was well ahead of his time” in contending that “[i]f two different goods can command an identical monetary price, then it can be assumed that they generate the same quantity of utility for the purchaser” (p. 25). “When Bentham idly wondered whether pulse rate or money might be the best measure of utility”, concludes Davies, “he could scarcely have imagined the industries that would develop dedicated to asserting and reinforcing the authority of particular indicators to represent our inner feelings” (p. 39).

However, it would not be until the mid-eighteenth century before there was a systematic attempt to construct quantitative measures of sensation. Davies cites the work of German “theologian-cum-physicist” (p. 27) Gustav Fechner, whose “psycho-physics”, which argued that “mind and matter are separate entities but must nevertheless have some stable, mathematical relationship to one another” (p. 29), as providing the breakthrough. For Davies, Fechner’s representation of “the relationship between mind and world as a numerical ratio”, “pointed the way to a more intimate micromanagement of individuals” (p. 35). The psychiatrists, therapists, and analysts who followed Fechner, turned their attention to “the subject having the feelings, rather than the object that seemed to be causing them” (p. 35).

Chapter two, “The Price of Pleasure”, explores the idea of an “equivalence between the sensations produced via the nervous system, and money” (p. 46). Davies argues that it is the work of the English psychologist, William Stanley Jevons, namely his Natural elements of political economy (1855), that firmly establishes “that economics could not ignore psychology any longer” (p. 50). “Given that labour was central to the classical economic view of capitalism”, writes Davies, Jevons’s insight was that “it must surely be relevant that workers suffer different
levels of pain as they go through their day, which then influences how much they are able to produce” (p. 50).

Jevons strove to found economics on a science of pleasure and pain, and like Bentham, regarded the mind as a kind of mathematical calculator. He is also “one of the architects of what is often referred to as _homo economicus_”, says Davies, “a somewhat miserable vision of a human being who is constantly calculating, putting prices on things, neurotically pursuing his own personal interests at every turn” (p. 61). For Jevons, “[t]o satisfy our wants to the utmost with the least effort—to procure the greatest amount of what is desirable at the expense of the least that is undesirable—in other words, to maximize pleasure, is the problem of economics” (p. 55). As such, for Jevons, the market “was a vast psychological audit, discovering and representing the desires of society” (p. 57). “Jevons was effectively turning the market into one vast mind-reading device”, comments Davies, “with prices—that is, money—as the instrument that made this possible” (p. 57). “The ideal of bringing the invisible realm of emotions and desires into the open”, he continues, “was now bound up with the ideal of the free market” (p. 57).

If the early part of the chapter shows how economics came to be connected with psychology through the work of Jevons, then the latter part shows how it came to be disconnected through the work of economists such as Alfred Marshall and Vilfredo Pareto, who introduced a theory of “preferences” in place of Jevons’s theory that “each pleasure and pain has its own discernible quantity” (p. 61). And by 1930, notes Davies, “the divorce of economics from psychology was complete” (p. 61).

With chapter three, “In the Mood to Buy”, comes a direct attack by Davies on the profession of psychology, particularly as it has developed in the United States. For it is in this chapter that Davies lays out how the history of psychology and the history of consumerism are intertwined projects, if not also co-dependent ones (p. 76). In this chapter, Davies argues that American psychology has “no philosophical heritage” and “was born into a world of big business and rapid social change, which risked spiraling out of control” (p. 85). He builds his case by showing how the first-ever psychology lab of the German experimental psychologist Wilhelm Wundt became the destination for many American psychologists in the late nineteenth century, but why it never became the model for the development of psychology in America. Rather, psychology in America was quickly co-opted by business interests who
were “increasingly hungry for knowledge they could use, especially regarding consumers” (p. 82).

The target in this chapter is less the experimental psychology of American psychologist-philosopher, William James, than the behavioral psychology of John B. Watson, which Davies says “was not merely anti-philosophical”, but also “virtually anti-psychological” (p. 89). For Davies, Watson’s “[b]ehaviourism stretches Bentham’s dream of a scientific politics to its limit, imagining that beneath the illusion of individual freedom lie the cold mechanics of cause and effect, observable only to the expert eye” (p. 88). Unlike Wundt, who conducted experiments on subjects who knew what was being tested, Watson did the opposite. For him, “subjects must remain partly ignorant of exactly what is being tested, or else there is the fear that they might adjust their behaviour accordingly” (p. 92). Watson’s greatest sin though was not the anti-philosophy or anti-psychology of his behavioralism, but rather that he sold it to Madison Avenue.

By 1920, the advertising industry was aware of the potential of using psychology to increase their effectiveness. And it was in the same year that Watson joined J. Walter Thompson, a large Madison Avenue advertising firm, at “a salary four times what he was earning at Johns Hopkins” (p. 94). Many other intellectuals though were also pulled into Madison Ave. Even Frankfurt School Marxist Theodor Adorno got into the act a short time later working on a study of CBS radio audiences (p. 99). During this period of merging the world of psychology with the world of business, the market was designed as a “space in which desires can be pursued but never fully satisfied, or else the hunger for consumption would dwindle” (p. 103).

Chapter four, “The Psychosomatic Worker”, brings the world of psychology even deeper into the world of capitalism by showing how and why managers and policy makers came to “yearn for a science of workplace happiness” (p. 109). According to Davies, however, “it was with that sort of hard science that many of our problems begin” (p. 109), namely, the birth of what he terms the “happiness industry”. Writes Davies:

For Bentham, happiness was something which resulted from certain activities and choices. Neo-classical economists such as Jevons and behaviourist psychologists such as Watson assumed something similar, implying that individuals could be lured to make certain choices by dangling a pleasurable carrot in front of them. But in the
context of business consultancy and individual coaching, happiness looks altogether different. Suddenly, it is represented as an input to certain strategies and projects, a resource to be drawn upon, which will yield more money in return. Bentham and Jevons’s psychological premise, that money yields a proportionate quantity of happiness, is spun on its head, suggesting instead that a quantity of happiness will yield a certain amount of money (p. 114).

Symptomatic of this reversal, Davies cites neuroscientist Paul Zak, who suggests that we view happiness like a muscle that requires regular exercising in order to keep it strong and healthy (p. 114). This chapter does a fine job in showing how our notion of a “good worker” has come full-circle since the 1870s when Frederick Winslow Taylor was developing his brutalist approach to management (the notorious “Taylorism” that would be laid out in The principles of scientific management (1911)) to today’s scientific scrutiny of bodies, movement, and performance in the workplace. Davies ends this chapter asking whether or not the latter are a “discreet return of the ‘scientific management’” of Taylor (p. 136).

It is finally in chapter 5, “The Crisis of Authority” that the ugly head of neoliberalism as a contemporary problem finally reveals itself. For Davies, neoliberalism is a “depressive-competitive disorder” that “arises because the injunction to achieve a higher utility score—be that measured in money or physical symptoms—becomes privatized” (p. 179). He shows how the Chicago school of economics (and the St. Louis School of psychiatry) breaks with the logic that says we have a moral and political responsibility toward the weak—one that often asks us to impose restrictions on the strong. For Davies, authority in neoliberalism “consists simply in measuring, rating, comparing and contrasting the strong and the weak without judgment, showing the weak how much stronger they might be, and confirming to the strong that they are winning, at least for the time being” (p. 179). Or, in short, that the very rich, successful, and healthy firms and people should become even more so.

This chapter, one of the best in the book, explains both how Thatcher and Reagan ushered in the era of neoliberalism (the more well-known story) as well as the role of a “renewed reverence for both competitiveness and the management of happiness” in its rise (the less well-known story). Davies is careful to note that American neoliberalism does not favor competitive markets, but rather markets as “a space for victors to achieve ever-greater glory and exploit the spoils” (p. 160).
other words, competition according to the Chicago school of economics was about destroying rivals, not co-existing with them—and the market was the central site of this destruction. The key shift in the chapter though is from an account of the neoliberal man who is “possessed with egoism, aggression, and optimism of a Milton Friedman or a Steve Jobs” (p. 161)—to the many others who are not.

It is at this point that Davies turns his attention to the development of anti-depressant drugs such as Prozac and to changes in the various editions of the *Diagnostic and statistical manual of mental disorders* (DSM). In 1980, DSM-III was published—a text which Davies describes as one of the most controversial and revolutionary in the history of psychiatry (p. 174). Whereas its predecessor (DSM-II) had 180 categories over 134 pages, DSM-III had 292 categories over 597 pages; whereas DSM-II said a symptom had to be present for a month before diagnosis was possible, DSM-III reduced it to two weeks; and most importantly, with DSM-III mental illness was detectable by observation and classification without need for an explanation why it had arisen (p. 174). According to Davies, DSM-III replaces “psychiatric insight into the recesses and conflicts of the human self” and the psychiatrists who conduct psychiatric counseling with “a dispassionate, scientific guide for naming symptoms”, which can be utilized for prescriptions by medical doctors and primary care practitioners (p. 174). In short, with DSM-III, psychiatry was cut out of the happiness industry to the point where today 80% of the prescriptions for antidepressant drugs in the US are written by medical doctors and primary care practitioners, and not by psychiatrists (p. 175).

Chapter 6, “Social Optimization”, takes the destruction wrought by the neoliberal era from the individual to the group. It shows how the interweaving of the science of happiness with social media innovation in the age of neoliberalism brings about its own unique set of problems. For example, it has been shown that the social media technology, Facebook, actually makes people feel worse about their lives rather than better. “If happiness resides in discovering relationships which are less ego-oriented, less purely hedonistic, than those which an individualistic society offers”, writes Davies, “then Facebook and similar forms of social media are rarely recipes for happiness” (p. 209). “The depressed and the lonely, who have entered the purview of policy-making now that their problems have become visible to doctors and neuroscientists”,

---

*THE HAPPINESS INDUSTRY / BOOK REVIEW*
concludes Davies, “exhibit much that has gone wrong under the neoliberal model of capitalism” (p. 211).

Davies describes a world where individuals seeking to “escape relentless self-reliance and self-reflection” (p. 211) turn to social media such as Facebook only to find that it further deepens the malaise brought upon them by the extreme individualism of neoliberalism. In short, “neoliberal socialism” sees the “social” as “an instrument for one’s own medical, emotional or monetary gain”—and in doing so perpetuates the “vicious circle of self-reflection and self-improvement” (p. 212). “Once social relationships can be viewed as medical and biological properties of the human body”, writes Davies, “they can become dragged into the limitless pursuit of self-optimization that counts for happiness in the age of neoliberalism” (p. 213).

The penultimate chapter, “Living in the Lab”, takes a look at the explosion of happiness and wellness data, which Davies views as “an effect of new technologies and practices of surveillance” (p. 219). The rise of “big data” is different from that of the “data survey”, as the latter was collected with the intention of analyzing—whereas with the former this is not the case (p. 233). Davies speculates that the “dream that pushes ‘data science’ forwards is that we might one day be able to dispense with separate disciplines of economics, psychology, sociology, management and so on”, replacing it instead with a “general science of choice” (p. 237). For him, “‘the end of theory’ means the end of parallel disciplines, and a dawning era in which neuroscience and big data analytics are synthesized into a set of hard laws of decision-making” (p. 237). “Add mass behavioral surveillance to neuroscience”, continues Davies, “and you have a cottage industry of decision experts, ready to predict how an individual will behave under different circumstances” (p. 238). In an ironic twist, the history of consumerism may even soon be looking at “predictive shopping”, where our purchasing decisions are made for us based on algorithmic analysis or smart-home monitoring—and sent directly to our home without our having to “ask” for them (p. 239).

In the final chapter, “Critical Animals”, Davies looks for an alternative to the Benthamite and behaviorist traditions critiqued throughout the book, that is, those “which view psychology as a step towards physiology and/or economics, precisely so as to shut the door on politics” (p. 267). What if, in contradistinction to these traditions, we view psychology as a “door through which we pass on the way to
“political dialogue” (p. 267)? In one of my favorite lines from the book, one which supports his call for political dialogue, he writes “‘Critique’ will not show up in the brain, which is not to say that nothing happens at a neurological level when we exercise critical judgement” (p. 269). “The attempt to drag all forms of negativity under a single neural or mental definition of unhappiness (often classified as depression)”, notes Davies, “is perhaps the most pernicious of the political consequences of utilitarianism in general” (p. 269). Pernicious, yes. But also yet another explanation as to why some today feel that critique has lost a lot of its steam (see, for example, Di Leo 2014).

Davies concludes that if only we took a fraction of the billions of dollars we spend “monitoring, predicting, treating, visualizing, anticipating the smallest vagaries of our minds, feelings and brains” and spent it “instead on designing and implementing alternative forms of political-economic organization”, we would at least start to move in a better direction (p. 271). But fat chance, right? Still, Davies notes that inroads are being made to de-medicalize misery albeit in direct opposition to the pharmaceutical industry and its representatives within the American Psychiatric Association (p. 271).

Critique and resistance though to the joining up of economics and medicine will not come easily. The science of well-being and “the monistic fantasy of a single measure of human optimality” is firmly entrenched in the age of neoliberal managerialism and measurement (p. 274). Davies says that we should consider adopting as a point of principle that “the pursuit of health and the pursuit of money should remain in entirely separate sphere” (p. 274). On the final page of the book, Davies asks, “What would a critique of smartness look like?” (p. 276). To be sure, from the perspective of The happiness industry, it is difficult to see what it would be. Davies’s historical account of the collaboration among psychology, economics, and business from the age of Bentham to the era of neoliberalism is convincing—and sets the stage for only deeper affiliations, if not the general science of choice noted above.

While historians of psychology and economics will surely want to haggle with some of the details of Davies’s accounts of these respective areas, these disagreements should not get in the way of appreciating his amazing account of “how government and big business sold us well-being”. The value of this book is the way in which it links developments in economics to those in psychology and social and political philosophy,
particular in drawing these into a dialogue with the work of Bentham, Friedman, Watson, and many others. The happiness industry is a thought-provoking and daring intervention into the crowded field of neoliberal political economy. Nevertheless, its bold theses and elegant historical foundation provides political economists with much new material to consider as the object of critique.

REFERENCES

Contact e-mail: <dileoj@uhv.edu>

N. Emrah Aydinonat  
*Bahçeşehir University*

In this review, I focus on issues that will interest philosophers of economics (for a more general review and a chapter by chapter guide to *Economics rules*, see Aydinonat 2015).

There is a cartoon that pictures a man in bed with an economic model. It reads: “economists do it with models!” Yes, in fact they do it with models. However, it is not entirely clear what they do with them and how. Dani Rodrik’s *Economics rules* is an attempt to explain what good economists do with their models and maybe more importantly what they should not do with them.

*Economics rules* is a work in philosophy of economics, written by a prominent economist. When economists enter the domain of philosophy of economics, they commonly overlook the work of professional philosophers of economics. It is of course possible that they are not aware of the work of philosophers, but an equally likely reason for their lack of consideration is that they (probably) think that they have nothing to learn from philosophers concerning the nature of economics and economic models. They are economists after all! Nevertheless, many economists fail to give a satisfactory answer to an important question: How can economic models that contain highly unrealistic assumptions help us in explaining (and understanding) real world economic phenomena? This question is more difficult than it appears and it does not lend itself to a quick and easy answer. For example, arguing that the real world is complex and that we need simplifications does not constitute a full answer, neither does using the map analogy (i.e., models are like maps; they are useful because they simplify!)

Dani Rodrik’s *Economics rules* is distinct in its attempt to provide us with a full and satisfactory answer to the aforementioned question. True, he also uses the map analogy, but he does not leave it at that. Rather, he discusses the nature and significance of economic models in detail. Providing a wide variety of examples from economic theory and
policy, he develops a new account of models and argues that the diversity of economic models is the key to solving the puzzle concerning the epistemic role of highly abstract models in economics. And a pleasant surprise for my fellow philosophers of economists: Rodrik—wait for it—utilizes philosophy of economics in his account of economic models.

Rodrik’s main aim in the book is to “explain why economics sometimes gets it right and sometimes doesn’t” (p. 5). On the one hand, Rodrik defends economics against popular lines of criticism of economics; on the other, he criticizes some of the practices in economics. More appropriately, he defends economics against those people—economists as well as non-economists—who, according to Rodrik, misinterpret economic models. Doing this requires an account of models that can obliterate misunderstandings shared by some economists and many critics of economics. In Rodrik’s account, common misunderstandings stem from the fact that both economists and non-economists sometimes mistake a model, with the model: They overlook the fact that economics is a collection of models and assume that the model they happen to be confronted with is the only model concerning the question at hand. It is for this reason that critics fail to see how unrealistic models can contribute to our understanding. And for the same reason, economists (and policy makers) who think that the model at hand will give them all the answers concerning the real-world problem they are facing are wrong. “It is a model, not the model” is the motto of Economics rules. In order to fully understand what it means, we need to take a look at Rodrik’s account of models.

Rodrik’s account of models is similar to Uskali Mäki’s account in some respects. Rodrik believes that economists use the method of isolation (Mäki 1992; Mäki 2010) and echoes Mäki (2005) in arguing that models are similar to experiments:¹

Many assumptions that go into economic models—perfect competition, perfect information, perfect foresight—are patently untrue. But […] models with unrealistic assumptions can be as useful as lab experiments performed under conditions that depart starkly from the real world. Both allow us to identify a cause-effect

¹ The reader may be wondering whether Rodrik cites Uskali Mäki. Yes, he does. He also cites Nancy Cartwright’s work. Note however that the other references in this review are not cited by Rodrik.
relationship by isolating it from other confounding factors (p. 180, emphasis added).

Thus, Rodrik contends that even though models misrepresent reality in many respects, they isolate real causal mechanisms. Or as he puts it “models are never true; but there is truth in models” (p. 44)—again echoing Uskali Mäki (2011). Rodrik takes it that isolating causal mechanisms and studying them (under the conditions specified by the model) helps economists in learning about “tendencies” and “likely consequences” (p. 45). Nevertheless, the isolation account of models does not fully answer how economic models that utilize unrealistic isolating assumptions (and unrealistic assumptions concerning background conditions) may help us in providing correct explanations or in understanding economic phenomena. Rodrik is fully aware of this. In fact, he is attentive to Nancy Cartwright’s (2007) warning that it is difficult to “truly isolate cause and effect in economics” (p. 44). In order to solve the puzzle presented by unrealistic economic models, Rodrik amends Mäki’s account with an additional observation: He argues that it is the diversity and the multiplicity of economic models that make economics a powerful social science. Note here that neither Mäki nor other proponents of the isolation view reject the diversity of models, it is just that they do not put enough emphasis on it.

Rodrik argues that having multiple models enables economists “to alter the background conditions selectively, to ascertain which, if any, make a substantive contribution to the effect” (p. 44). It may first seem that Rodrik is talking about the robustness of economic models. However, he takes it that non-robust models are also valuable because they teach us about the conditions under which their results would not hold (cf. Kuorikoski, et al. 2010). If a model’s results change under slightly different conditions, we learn that the results of the model are context-specific. In fact, Rodrik argues that “the correct answer to almost any question in economics is: It depends” (p.16). But “models do more than warn us that results could go either way. They are useful because they tell us precisely what the likely outcomes depend on” (p. 17). Thus we may interpret Rodrik as follows: It is not exactly true to say that models teach us about how causal mechanisms work in isolation; rather models teach us about how ‘structure’ (conditions specified by the model) affect the outcome of the isolated (set of) causal mechanisms (cf. Cartwright 2009; Aydinonat 2008; Aydinonat 2007).
The emphasis on the diversity of economic models is important because viewing economics as a collection of models facilitates a better understanding of how highly abstract economic models help us explain. Rodrik’s account of models has similarities with the *cluster view of models* (Ylikoski and Aydinonat 2014). Different models are connected to each other in many ways; they may isolate different causal mechanisms and play different roles in the cluster of models that are relevant to the phenomenon at stake. If one fails to see this, Rodrik believes, one fails to see the true value of economic models. Different models provide us with different and sometimes conflicting answers. This diversity of models is valuable because the collection of these models “enable the accumulation of knowledge, by *expanding the set of plausible explanations* for, and our understanding of, a variety of social phenomena” (p. 46, emphasis added). Each model may seem limited in what it can accomplish, however each of these models increase our ability to explain by way of expanding our *menu of possible explanations* (see Ylikoski and Aydinonat 2014)—so that “we have *a menu to choose from*” (pp. 73-74, emphasis added).

Knowledge accumulates in economics not vertically, with better models replacing worse ones, but horizontally, with newer models explaining aspects of social outcomes that were unaddressed earlier. Fresh models don’t really replace older ones. They bring in a new dimension that may be more relevant in some settings (p. 67).

If economics is a powerful science it is because economics is a collection of a wide variety of models that can help us explain a wide variety of economic phenomena. Rodrik claims that “just as social reality admits a wide range of possibilities, economic models alert us to a variety of scenarios” (p. 209). So, economic models “are contextual and come in almost infinite variety” (p. 114). In Rodrik’s account, viewing, using or criticizing economic models in isolation from other related models is an important mistake. Similarly, disregarding the conditions under which the model results hold is a big mistake. Since these are common mistakes, what I call the cluster view of economic models helps economists defend economics against many types of criticism. In the book, Rodrik contents that the popular lines of criticism fail because they ignore the variety and multiplicity of models in economics. Critics commonly cite the representational properties of a single model and conclude that it is too unrealistic to be true. Similarly, some economists...
think that their favored model is *the* model, and assume that it will give them all the answers. However, as Rodrik argues, a single model would only give us a partial answer: “They [models] provide at best *partial explanations*, and they claim to be no more than abstractions designed to clarify particular mechanisms of interaction and causal channels” (p. 114; see Aydinonat 2007, 2008 for a similar argument). In order to be able to provide satisfactory singular explanations of economic phenomena economists ordinarily utilize a set of models and try to select the right model—or, the right combination of models—for the explanatory task at hand. The isolation view helps us understand how one model is related to the real world, the cluster view helps us understand how economists explain and how economic models contribute to our understanding. In *Economics rules*, Rodrik illustrates this with many examples from economics.

Rodrik takes it that despite their unrealisticness, there should be some grain of truth in economic models. But where does truth reside in models? According to Rodrik, the explanatory power of economic models depends on the realisticness of their *critical assumptions*. Although it is not entirely clear which assumptions of a model are critical, the examples Rodrik provides imply that they are assumptions relating to applicability (i.e., helping economists decide whether the model can be utilized in explaining the particular case at hand). For example, a model that assumes perfect competition cannot explain cases of imperfect competition. Or, a model that assumes that firms have market power cannot be used to explain cases where firms do not have market power. Critical assumptions are important in using and applying models. Moreover, we understand from Rodrik’s account that the isolated causal mechanism should represent a real causal mechanism in order to be explanatory. If the proposed mechanism is nonexistent, the model cannot explain the case at hand. So, at first glance it seems that truth resides in the isolated causal mechanism and the truth of the model’s critical assumptions is important for its applicability. Nevertheless, Rodrik also argues that one of economic models’ important contributions is to “open our eyes to counterintuitive possibilities, and unexpected consequences” (p. 46). True, models help us find out that known causal mechanisms may produce unexpected, counterintuitive results. But models also help us discover new causal mechanisms and explanatory factors. This should be one of the ways in which they help us expand our menu of plausible explanations.
However, if the results of a model depend on its assumptions, it would be hard to establish that it helps us in discovering new causal mechanisms operating in the real world. This is because, it would be difficult to ascertain where the model’s results are coming from: Is it the result of its (unrealistic) assumptions or, of the proposed causal mechanism? Remember that Rodrik takes it that the true answer to almost any question in economics is “it depends”. The difficulty is to know what the result depends on. With so many unrealistic assumptions concerning functional forms, mathematical properties, background conditions, rationality, and so on, how can we ever know what the model is teaching us about the real world?

Rodrik’s answer to these questions is that of a practicing economist. He lists four different (but not exclusive) “verification” strategies that would help us in selecting the right model. The first one is verifying models’ critical assumptions and ascertaining that the model is representative of and applicable to the case at stake. The second strategy is verifying the existence and operation of the proposed mechanisms. The third strategy concerns models that are built up from basic principles: Check whether the results of the model hold in the real world. And the fourth strategy is to check whether other implications of the model are consistent with our observations. Although this broad list of principles may not satisfy philosophers, Rodrik gives examples for each strategy and demonstrates that developing and using models is a craft—requiring economic intuition, experience, etc.—and that it cannot have precise rules: “Good judgement and experience are indispensable, and training can get you only so far” (p. 83). Moreover, it all depends on the task at hand. A particular model may be useful for some explanatory tasks and not for others. The purpose and limitations of a model, as well as the aims of the economist using the model are important elements in the verification process. So the true answer to questions concerning the locus of truth and critical assumptions is: it depends. Rodrik argues that “the key skill is being able to move back and forth between candidate models and real world” (p. 93). Accordingly, we can know whether a model helps us in discovering new causal mechanisms only after verifying that the proposed causal mechanism exists in the real world. Once again, the proof of the pudding seems to be in the eating.

Rodrik’s guidelines for “verification” are definitely useful and enlightening, particularly because he gives several examples from
economics that helps the reader get a feeling of how the “verification” process works. However, these guidelines are somewhat vague for philosophers’ taste. Thus, there is room for improvement and hence an invitation for philosophers of economics right here. We can reach a better understanding of the art and craft of verification by way of studying the process of model development and the way in which models are utilized in providing explanations. The precision in defining critical assumptions and their verification can be improved. Moreover, we need to improve our understanding of how isolation works, what it achieves, and how it relates to the aims of modelers and practitioners. Finally, if it is the case that the diversity of models is what makes economics a powerful science—and I believe it is the case—we need a better account of how models relate to each other and how their collection relates to the real world. It is time for philosophers of economists to engage in the world of models and start appreciating their multiplicity and diversity. We shall not forget: It is a model, not the model!

In sum, Economics rules is an excellent book on the nature of economics. It is rich with examples from history of economics and current debates in economics. It is a must read for all students, practitioners and philosophers of economics; and an indispensable guide for anyone who wishes to have a better understanding of economics and how economic models work. Economics rules also contains much to think and disagree about for philosophers of economics. It will spark new debates and more fine grained work in philosophy of economics. I know that the readers of this review are wondering why I did not present my own criticism of Rodrik’s account. The reason is that I wanted to use this limited space to show that it is a very interesting book for philosophers of economics. In fact, my aim was to show that philosophers of economics can only fail to read it at their own peril. I am sure that we will have plenty of opportunities to discuss Rodrik’s account of models. Let us get down to the nitty-gritties of Economics rules later on.

REFERENCES


**N. Emrah Aydınonat** is an associate professor of economics at Bahçeşehir University (Istanbul, Turkey). His research focuses on philosophy of economics, in particular on models and explanation in economics. He is the author of *The invisible hand in economics* (Routledge, 2008) and the co-editor of *Economics made fun: philosophy of the pop-economics* (with Jack Vromen, Routledge, 2015).

For more information: [http://neaydinonat.com](http://neaydinonat.com)

Contact e-mail: <aydinonat@gmail.com>

**Alessandra Basso**

*University of Helsinki*

*Science outside the laboratory* brings the reader along an exploration into the epistemology of scientific investigation in the field. The book is organized in such a way as to move from general epistemological questions to more specific methodological issues, with each chapter diving deeper into a specific problem related to the previous chapters. In this review, I focus on a line of argument which I think cuts across the chapters and allows me to highlight the main contributions of the book.

Starting from the identification of the specific problems one faces when conducting scientific research in the field, *Science outside the laboratory* sets up to provide a positive answer to the question of whether it is possible to obtain reliable knowledge when laboratory investigation is not possible. In order to do so, Boumans provides an account of measurement of field phenomena and addresses methodological issues related to its validation. From this discussion, it emerges that the validation of field investigation requires field-specific assessment methods and, moreover, that expert judgment is needed in order to make reliable field investigation. The book, therefore, moves on to discuss what kind of judgment is needed, and what methods can be employed to validate experts judgments in the context of field investigation. This nicely closes the book’s narrative circle and encourages the reader to look back at the opening question with an optimistic attitude: observation and measurement outside the laboratory can be rigorous, but it requires field-specific criteria for evaluating epistemic value. The suggestions offered for developing assessment methods that can be used to validate field investigation are, I believe, among the book’s main contributions.

**RELIABLE SCIENCE OUTSIDE THE LABORATORY**

The methodological investigation offered in *Science outside the laboratory* is motivated by the attempt to deal with some of the
problems that arise when investigating phenomena that, for practical, technical or ethical reasons, cannot be studied in the laboratory. The book uses the term ‘field science’ to refer to the variety of research practices outside the laboratory, which studies phenomena that cannot be isolated from their environment and hence cannot be investigated by means of manipulation or intervention (p. 2). Field science, therefore, seems to be a category of scientific practice that intersects the natural and the social sciences. However, some of the considerations made in the book appear to be related more to the opposition between natural and social sciences than to the contrast between field and laboratory science.

One aspect of the book where the issues at hand seem to cut across the social science/natural science divide rather than the field/laboratory distinction is the discussion of the challenges that emerge when conducting scientific research outside the laboratory. Boumans highlights in particular three problems, but only one of them appears to be specific to field science. First, a specific problem of field science is that, in contrast to the results of laboratory experiments, field results are not reproducible, and hence they are less objective, because they rely more on the credibility of the observer. Second, the gathered data is often based on evasive answers, lies or misunderstandings and hence research practice has to take into account the possibility of deliberate deceit from the individuals under investigation. This problem is not specific to field science: it does not arise when conducting field investigation in the physical sciences and, instead, it may come about in the social and behavioral sciences, whether conducted in the field or not. Nature does not lie, but individuals may well do so even if they are in the controlled environment of laboratory experiments.

Third, Boumans argues that field science is more inexact than laboratory science, leading to more incomplete theories and models (p. 3). As an example, he provides a thorough discussion of the difficulties related to the selection of the right set of variables to be included in economic models. Based also on the work of Trygve Haavelmo, Boumans argues that selecting the right set of variables is especially difficult due to the inexactness of economic theory. It seems to me that this kind of problem is not specific to field science. It may affect both the natural and the social sciences, although arguably the social sciences are more commonly affected by it. The incompleteness of theories and models in certain areas of science, however, might influence both laboratory and
field investigation in these contexts. For instance, it might be argued that rational choice theory is unable to guide the choice of the right set of variables to be included in models of field phenomena like the business cycle, as well as in models of laboratory phenomena like preference change.

The intersection between the categories of field science and social science can be seen as providing potentially interesting perspectives for investigating the epistemology of field science and its relation to the debates regarding the social sciences. On the one hand, Boumans’s focus on field science can offer a somehow new perspective to address some of the problems that affect the social sciences, and, on the other hand, some of his conclusions can be interpreted as contributing to general debates regarding the epistemology of the social sciences. However, to fully exploit this potential, it would have been interesting to see a clearer distinction between these two categories of scientific practice and a discussion of the relations between them.

The philosophical debates concerning the reliability of social scientific investigation are not foreign to social scientists themselves. As an example of how these problems are examined within social scientific disciplines, Boumans discusses Oskar Morgenstern’s considerations about how economic observation is different from observation in the natural sciences. Two of Morgenstern’s ideas are particularly relevant. First, the common assumption made in statistical regressions that errors average out over a large number of observations is not warranted for economic data, and neither is there good evidence to support it. Due to the specific nature of economic observations, and more generally of social phenomena, divergence in the distribution of data is to be expected, and hence one cannot assume that the data is normally distributed. Second, Morgenstern argues that scientific observation should be guided by theory, but economic theory is highly inexact and hence cannot adequately fulfill this task.

Morgenstern’s analysis has a pessimistic tone and ends up being skeptical about the possibility of making reliable observations in economics. Boumans does not share this pessimism. For him, Morgenstern’s skeptical conclusions are brought about by the economist’s propensity to take the natural sciences as the gold standard of reliable observation, the reference against which economic observation should be compared. Boumans’s analysis, instead, is based on the idea that the criteria for evaluating the reliability of field science
should be local: the methods of field investigation should obey the standards of social science (p. 16).

AN ACCOUNT OF MEASUREMENT
In order to explore the methods that can be used to validate field investigation, Boumans begins by providing an account of measurement that can deal with the peculiar characteristics of field science. The account he develops is based on a prominent theoretical approach, the representational theory of measurement, and on the authoritative methodological guidance offered by current metrology. Since neither one is satisfactory when dealing with measurement of field phenomena, Boumans extends and adjusts them so as to arrive at a representational account of measurement that is consistent with inexact theory and incomplete models.

Representational accounts are based on the idea that measurement outcomes represent certain qualitative features of the system under measurement. The crucial question that these accounts have to address is what underpins the claim that the outcomes represent the measured property. The representational theory of measurement developed by Krantz, Luce, Suppes, and Tversky in the three-volume *Foundations of measurement*, takes an axiomatic approach to answer this question. The representational relation is established by providing a set of formal axioms about the structure of the property under measurement and a logical proof that, under these assumptions, the measurement outcomes are a homomorphic mapping of the qualitative structure of that property.

A criticism often raised against the axiomatic approach is that it lacks empirical bases, because the assumptions about the property under measurement are not tested empirically. To discuss this issue, *Science outside the laboratory* provides a brief historical excursus into the development of the axiomatic approach, which throws light on the progressive shift from an empirical understanding of the axioms to the essentially formal perspective of the representational theory of measurement. With Norman R. Campbell, and partially also with Stanley Smith Stevens, measurement is interpreted as being grounded on certain characteristics of the property of interest, like order and additivity, which are understood as hypotheses about facts of the world that have to be tested and proven empirically. With Suppes and coauthors, instead, the foundations of measurement are established under the
condition that the empirical system has certain characteristics, without bothering with the empirical justification of these assumptions.

Although I agree with Boumans that the historical development of the axiomatic approach reveals a shift from an empirical to a formal understanding of the axioms, it should be noticed that, apart from relatively simple cases like the measurement of length, providing an empirical interpretation of the axioms does not necessarily mean that they are in fact tested empirically. Testing whether certain quantities, like loudness or happiness, satisfy the assumption of additivity, for instance, can be challenging or even impossible. Indeed, realist representational accounts of measurement, which argue for an empirical understanding of the axioms, maintain that there is little point in testing the axioms one by one, because a measurement is not required to rely only on true assumptions about a phenomenon in order for its outcomes to provide some true information about that phenomenon (Swoyer 1987).

Boumans argues that the axiomatic approach is not suitable for field measurement, independently of whether the axioms are tested or simply assumed. In his view, homomorphic representation requires a white-box model based on complete knowledge of the property under measurement. However, complete knowledge is not possible in field science, and hence a representational account of field measurement cannot be based on homomorphism. As an alternative, Boumans suggests extending the representational theory of measurement so as to encompass also cases in which one can aspire only to gray-box models, that is, modular models where the modules are black-boxes (p. 50).

Since gray-box models do not allow establishing homomorphism, the question of what grounds the claim that the outcomes are reliable representations should be addressed in a different way. Boumans's approach to answer this question is by looking at the practical guidance for the assessment of measurement reliability offered by current metrology.

In current metrology, the assessment of measurement is based on the evaluation of uncertainty. This is done partially by means of statistical methods and partially by other means, like skilled judgments and calibration against externally accepted references. The statistical methods are based on the evaluation of outcomes' stability under controlled variations of the measurement conditions, and therefore, according to Boumans, they are made for exploiting the controlled
conditions of the laboratory. Since controlled variations might be impossible in the field, Boumans argues that the role of statistical methods in the assessment of field measurement should be narrower than in laboratory science. To compensate, non-statistical methods should have a broader role. Moreover, since field science lacks authoritative institutions able to fix standard references for calibration, the only available source of information for non-statistical assessments of field measurements is skilled judgment. As a consequence, the assessment of field measurement is more subjective than the one of laboratory measurement, but this does not mean that it is less rigorous. On the contrary, according to Boumans field measurement simply requires different assessment methods.

On Boumans’s account, an objective assessment of field measurement requires constructing a model of the measurement process, which should include all relevant factors suggested by theory, expert judgment and background knowledge, and, then, testing the validity of this encompassing model. Due to the inexactness of theory, this model might still be incomplete, but the validation test will tell whether a significant factor is missing. The validation of these gray-box models should be done by means of ‘behavior pattern tests’: instead of testing the validity of the model structure, the assessment is based on how accurately the model can reproduce the major patterns exhibited by the target phenomenon.

By emphasizing the role of models in underpinning claims about the epistemic reliability of measurement, Boumans’s suggestion contributes to a recent body of literature on model-based accounts of measurement (see Tal 2013, 2015). In particular, the book can be seen as defending the idea that theoretical models of the measurement process are necessary in order to obtain outcomes that can be considered as reliable representations of the properties under measurement. Different kinds of models can serve this purpose, ranging from detailed white-box analytical models to gray-box models based on inexact theory. I believe that model-based accounts of measurement can offer specific advantages in the context of field science, because they allow making claims about measurement reliability also in cases where it is not possible to control the actual system under measurement. However, it is worth noticing that the relevance of model-based accounts of measurement is not limited to field science. Models can play a role in
underwriting reliability claims also in the measurement of physical quantities in the laboratory (Tal 2011).

The philosophical discussion of the role of models in measurement practice is still at an early stage and many questions remain open, concerning for instance what kind of model is employed and how, as well as what connections can be made with the wide philosophical literature on scientific modeling. Therefore, I think that Boumans's model-based account could come to full life in the investigation of particular cases of successful measurement practice in the field. Although Science outside the laboratory is filled with historically detailed examples, which give an overview of the methodology of scientific practice in the field, it would have been useful if Boumans had provided a specific case study of measurement practice which could help to illuminate some of these issues.

THE NEED FOR EXPERTS AND THEIR ASSESSMENT

Drawing on a number of examples, Boumans argues that reliable measurement in field science always requires expert judgment. Building models, selecting the relevant variables, choosing between alternative models, and assessing them require theory as well as additional information that comes from expert judgment.

By focusing in particular on Haavelmo's works, the book provides a thorough analysis of a methodological debate in econometrics about the possibility of testing the completeness of models. Haavelmo identifies an important problem faced by econometricians when testing the significance of the causal factors to be included in their models. Econometric models are tested against specific bodies of data. If the residual errors are large, then it is inferred that something is wrong with the model. But, if the residuals are small, Haavelmo argues that one can only assume that the model includes the right set of variables, without being able to prove it. According to Haavelmo, in case of small residuals, it is impossible to discern between two alternative cases: either the model is complete, or the data set happens to come from a contingent situation in which certain determinant variables are constant or uninfluential, and hence remain latent. In other words, a variable might appear uninfluential simply because the variation it displays in the data set is too small to show its effective potential influence. As a consequence, it might happen that the test yields small residuals even if certain relevant causal factors are not included in the model.
This discussion sheds light on a problem that is of general interest for field science: because of the impossibility of enforcing an experimental design in the field, we are stuck with Nature’s experiments. A field science cannot be an inductive science by relying only on the data coming from contingent situations and statistical methods to analyze it. According to Haavelmo, if Nature’s experiments are insufficient to evaluate the potential significance of causal factors, theory may help. But, since econometric theory is inexact, additional sources of information are needed in order to choose the right set of variables. This leaves open the question of where this additional information comes from: although Haavelmo refers to this kind of knowledge, he never specifies what kind of knowledge this is. According to Boumans, this additional source of information is provided by expert judgment.

Since making inferences from observations to phenomena and building accurate models require experts, the reliability of field investigation depends, in part, on the reliability of expert judgment. *Science outside the laboratory* offers a perspective on how to validate expert judgments by drawing on methods developed for laboratory science and adapting them so as to fit the context of field science. To assess the reliability of field investigation, Boumans suggests, the experts should be validated in a similar way as the measurement models.

Boumans’s proposal is based on a method of reaching consensus among experts which was developed for applications to engineering problems. According to the Cooke method, rational consensus among experts is based on a weighted average of the individual expert judgments, where the weights are determined by testing the experts’ estimations against the known value of reference variables (p. 159). When attempting to apply this method to the social sciences, however, it might prove difficult to individuate suitable variables for the test. In the social sciences, the variables that enjoy wide consensus on their value are commonly well known to the experts, making it pointless to use them as references for the validation of their judgments.

As a solution to this problem, Boumans suggests to opt for a model-based forecasting test. In economics, he argues, there is much more consensus on the validity of specific empirical models than on the values of specific variables. Therefore, he suggests to test experts’ judgments against the predictions made by valid models. More
precisely, expert judgment should be evaluated by means of behavior pattern test, that is, not only against point predictions, but also against a broader range of predictions about behavioral patterns, like frequencies, trends, phases, etc. (p. 171). Experts who regularly make good predictions of behavioral patterns should be assigned higher weights in the aggregated judgment.

With this methodological suggestion, *Science outside the laboratory* concludes its exploration of the methodology of reliable field science with an optimistic attitude. Reliable investigation of field phenomena is possible, but it requires a context specific methodology of assessment. In particular, assessing the reliability of field measurement requires the validation of both the measurement models and the involved expert judgments. The proposals for field-specific methods of assessment for measurement and expert judgment are, I believe, among the main contributions of the book. Interestingly, in both cases, the recommended assessment method is model-based, which highlights a growing attention towards the multiple roles played by models in field scientific practice and, more generally, in the social sciences.

In conclusion, *Science outside the laboratory* is a recommended reading for philosophers or social scientists interested in scientific practice outside the laboratory, its challenges, its methodology, but also for anyone who is interested in understanding the role of models in underpinning reliability claims in the social sciences.

**REFERENCES**


**Alessandra Basso** is a PhD student at the Academy of Finland Center of Excellence in the Philosophy of the Social Sciences. Her thesis is intended as a contribution to the philosophical investigation of measurement practice, and pays particular attention to measurement in the social and behavioral sciences.

Contact e-mail: <alessandra.basso@helsinki.fi>

JAAKKO KUORIKOSKI
University of Helsinki

The purposes and limitations of all types of models used in economics, past and beyond. This is the rather ambitious topic of Lawrence Boland's new book. The book reviews standard microeconomics, macroeconomics, game theory (including evolutionary), experimental economics, and econometrics (both micro and macro). In its mission statement, the book is declared methodology with a small 'm', to be distinguished with a capitalized Methodology discussed by Philosophers. Methodology discussed by Philosophers deals with lofty issues of 'Realism', ‘Testing’, and 'Explanation', whereas methodologists are interested in, well, such things as the realism of assumptions, the testing of models, and explanation. In the end, the biggest drawback of the book ends up being a seemingly aesthetic aversion to completely imagined capital letters. It is clear that the book discusses the very same issues that have vexed philosophers of science and philosophers of economics for at least the last few decades, and arguably for centuries, yet it refuses to engage these resources. This is a regrettable, but perhaps inevitable, consequence of the limits of any one person trying to cover the whole of economics and the relevant philosophy of economics.

Boland begins with a thesis about a paradigm and generational shift in economic modeling around the beginning of the 1980s. Economists having their postgraduate studies in the 1980s or later no longer conceive models as being models of a theory, built in order to provide an instrument with which to empirically test that theory. Instead, model building began to be equated with theorizing itself. This shift is argued for in a concise, yet convincing manner. And this shift does not sit all that well with the Popperian spirit of Boland's Methodology (yes, with a capital ‘M’). Although the book does not attempt to systematically build a single argument or an encompassing philosophical position, the overall message emerging from the discussion seems to be a plea for
more interest in the realisticness and empirical testing of modeling assumptions.

The book first goes through the stylized history of the division between micro and macroeconomics. It dates the beginning of modern microeconomics to Marshall and proceeds from there to discuss the need for, and the plausible form of, microfoundations. Standard criticisms against the use of representative agent constructs and the sterility of general equilibrium theory are also raised. The biggest faults located by Boland in theoretical macro models based on equilibrium assumptions are their lack of dynamics and any idea of adjustment process for prices. There is nothing new in these accusations, but Boland raises the issues with a peculiar twist. It would not, in fact, be enough to endow the model with dynamics by having the endogenous variables be determined by their lagged values and additional exogenous variables. This would be merely a combination of exogenous shocks and a predetermined time path, not truly endogenous dynamics resulting from the actions of free agents. Approvingly quoting Hicks, Boland holds that economics must not only include time, but be in time.

The argument here is that the problem is not just in the lack of an empirically credible adjustment process, but in the fact that the conceptions of learning and expectation formation implicit in any proposed sketch of an adjustment process are flawed, because they have always been based on inductive reasoning. And inductive reasoning was demonstrated to be impossible by Hume. This foundational problem of the groundlessness of inductive learning is also raised with respect to agents within game-theoretical models: common knowledge of rationality either assumes a foundationally irrational inductive learning process or simply assumes away the whole ‘problem of induction’. Even more generally, all economic applications of game and decision theory are suspect to the extent they appeal to probabilities and thus conflate decision under risk and uncertainty.

Now, there is certainly nothing wrong with taking such a strong position with respect to the roles of probability and inductive reasoning. Boland’s deep mistrust of both notions is very much in line with his general Popperian outlook: there is no quantitative (probabilistic) measure of (inductive) evidential support and the only respectable use of probability concepts is as propensities of well-defined probability set-ups (such as games of chance). What is problematic is that these stances are not really argued for, but simply claimed to be the only rationally
acceptable positions. In this book Boland thus inadvertently manages to show how deep Methodological commitments have direct implications to model choice and testing, but at the same time refuses to acknowledge or really argue for the adopted big-M presuppositions. Inductive reasoning is bogus because of Hume and probabilities should not be trusted because the world is really not stochastic. I fully agree with Boland in that too much of game theory is motivated by purely mathematical and conceptual puzzles (such as the conceptual mire of knowledge and belief, unnecessarily covered at length also in this book) whereas too little attention is spared for methodological reflection on how these models are supposed to connect with economic reality. I am not just sure whether the arguments presented for this claim are the most convincing ones.

And what goes for reasoning within the models also goes for reasoning about models. It comes as no surprise that Boland’s Popperian outlook is not hospitable to Bayesian philosophy of science, but it would have been nicer to acknowledge its existence and even review some of its claimed advantages, rather than dismissing the very possibility of such an enterprise as if resting on a simple failure to understand elementary logic or as an illegitimate act of moving of the goal posts after failing to score.

Instead of inductive reasoning, Boland advocates a simple hypothetico-deductive picture of testing. But in contrast to Popper and his followers, he aims to actually overcome the Duhem-Quine (D-Q) thesis and the problem of underdetermination by providing an account of “logically adequate” model testing. Boland first makes an important distinction between testing a substantial theoretical assumption (a behavioral hypothesis) and testing a specification assumption linking the first kind of assumption to empirical data. The way to test a specific behavioral hypothesis in a logically adequate way is to construct a model with an alternative behavioral hypothesis but similar specification assumptions—a counter-example to the original hypothesis—and seeing which fits better. If the counter-example fits but the original does not, then the behavioral hypothesis is refuted. This certainly sounds sensible and fits well with the contrastive ideas about testing and confirmation, as well as with the ideas about testing specific modeling assumptions using robustness analysis and the like, currently vigorously discussed in philosophy of science. Nevertheless, it is obvious that this is not a logically foolproof solution to the D-Q
“problem”. There will still always remain logically possible alternative explanations for why one model fits whereas the other does not. The D-Q thesis simply does not admit of such a conclusive “solution”.

It is clear that we are now engaged in Methodology with a giant M. And it is a missed opportunity that the book does not discuss other currently viable Popperian views, such as the severe testing framework. This is all the more puzzling since Boland is happy to review (approvingly) the work of Aris Spanos on the importance of statistical misspecification testing, but either ignores or refuses to engage with the broader Methodological picture painted by him in collaboration with Deborah Mayo (see, e.g., essays in Mayo and Spanos 2010). Also much of the discussion on the way that empirical models confront the empirical world could have been much more illuminating if structured along the Bogen-Woodward data-phenomena-theory framework (Bogen and Woodward 1988). Now, the discussion on testing is conceptualized as that of testing theoretically motivated models directly with observations (even token ones!). Treating the data-phenomenon inference as a separate epistemic step would, in my mind, make more sense of much of empirical practice and is, arguably, also a more charitable interpretation of such diverse things as the casual appeal to stylized facts, the role of laboratory experimentation, and the LSE-methodology in macroeconometrics.

When it comes to econometrics, especially macroeconometrics, the book criticizes much of empirical practice as necessarily presupposing that human actions are predetermined by nature-given and immutable parameters. This is not conceived of only as an empirical shortcoming, but as some sort of conceptual folly or a symptom of a world-view antagonistic towards human autonomy. Yet again, the biggest shortcoming in the way the book covers the central methodological disputes around econometric modeling is that it ignores the advances in understanding causality and prediction versus accommodation achieved in philosophy of science in the last few decades. For example, Boland seems to conflate the (completely reasonable) worry that any estimated parameter values may not be stable across time with the question of whether deep parameters invariant under exogenous interventions can be uniquely identified given the available data and restrictions. Yet it is arguably the latter concern that limits the usefulness of completely atheoretical VAR models. And again, there is an opportunity missed in that Boland does not discuss the important work of Kevin Hoover on the
methodology of macroeconometrics—work (e.g., Hoover 2001; Hoover 2010) that surely should have qualified as small-m methodology given Hoover’s disciplinary affiliation.

In sum, the book’s breadth of scope is breathtaking and it is a treasure trove of useful references for further pursuing more specific questions. I certainly learned a lot reading this book. But in trying to cover every aspect of orthodox economic practice the book, perhaps inevitably, fails to really develop a comprehensive and coherent methodological view from which to evaluate or perhaps even offer guidance concerning the activities in question. This robs the book much of its reason d’être. It covers a bewildering array of topics, and there are a lot of astute observations about them, but it is not quite clear to what ultimate end. There is certainly a strong methodological view present in the book, but it is left mostly implicit and scantily argued for. The refusal to explicitly engage with issues that are perceived as too philosophical does not mean that methodological reflection could proceed entirely without commitments to such issues—it just means that the resulting overall picture ends up feeling somewhat idiosyncratic and lacking in foundations.

But perhaps the biggest omission arising from snubbing philosophy of science relates to the title: considering that the book is not just about the limitations of models, but also about their purposes, it is peculiarly silent about what is arguably the main purpose of most highly idealized theoretical models, namely explanation or “sense-making”. But delving into this territory would probably have meant discussing more Methodology with a capital M.

REFERENCES


Jaakko Kuorikoski is currently a professor in social and moral philosophy at the University of Helsinki and a research fellow at TINT
Academy of Finland Centre of Excellence in the Philosophy of the Social Sciences. He has published on explanation, models and simulations, mechanisms, and causality in, e.g., British Journal for the Philosophy of Science, Erkenntnis, Philosophy of Science, Philosophy and Phenomenological Research, and Synthese. His current research interests include evidence, macroeconomic modeling, and the social and policy relevance of neuroscience.

Contact: <http://jaakkokuorikoski.wordpress.com/>
PHD THESIS SUMMARY:
Reconciling normative and behavioural economics

GUILHEM LECOUTEUX
PhD in economics, May 2015
École Polytechnique

Economic analysis is traditionally separated into two different branches: positive economics seeks to explain and describe how economic agents behave, while normative economics aims at evaluating economic outcomes, policies, and institutions. Both approaches attribute to economic agents coherent preferences, i.e., preferences that are consistent and context-independent. Individual behaviour is described by assuming that agents behave as if seeking to satisfy their preferences, and economic outcomes are desirable to the extent that individual preferences are satisfied. Conventional normative economics therefore takes the satisfaction of individual preferences as the normative criterion: although this approach seems perfectly justifiable when individual preferences are coherent, it is less obvious that this criterion remains valid when individual preferences are incoherent, as suggested by the experimental findings of behavioural economics (it is indeed unclear that it is desirable to give to someone what she prefers now, knowing that she is likely to change her mind in other circumstances). The problem of how to reconcile normative and behavioural economics (the ‘reconciliation problem’, McQuillin and Sugden 2012) is that economists need to develop an alternative normative criterion to the standard preference satisfaction paradigm. The main approach suggested up to now—behavioural welfare economics (henceforth ‘BWE’)—consists in assuming that the revealed preferences of the agents, when they are incoherent, are a combination of true preferences (the preferences the individual would have revealed, were she perfectly rational) and errors. The satisfaction of the underlying true preferences is then taken as the normative criterion (see for instance Thaler and Sunstein’s (2008) libertarian paternalism and the use of nudges).

I argue in the first part of my thesis that this notion of true preferences is psychologically and philosophically problematic:
behavioural welfare economics indeed represents the individual as if it was an ‘inner rational agent trapped in an outer psychological shell’ (Infante, et al. 2015a, 2015b), since psychology is only conceived as a process that is likely to interfere with a latent ‘rational’ mode of reasoning. This mode of reasoning should be able to generate a unique coherent preference ordering (the agent’s true preferences), if the agent’s reasoning was not influenced by irrelevant psychological factors (chapter 1). I then trace the historical origins of the notion of true preferences, so as to understand why many behavioural economists accepted this notion as the primitive of their models. I argue that the model of the inner rational agent was already implicit in neoclassical economics from Pareto on, although it was historically designed as a model of representative agent, and therefore could not be used to infer any conclusion about individual welfare (chapter 2). I then highlight the methodological continuity between Pareto’s work and BWE: BWE indeed models the individual as the combination of a homo economicus and a homo psychologicus, but shares Pareto’s methodological difficulties in defining unambiguously the preferences of the homo economicus (chapter 3). I illustrate the methodological and philosophical difficulties of this reductionist approach in the case of intertemporal choices (Lecouteux 2015). I conclude the first part by questioning the ethical claim of BWE, according to which what matters is the satisfaction of one’s true preferences, i.e., of the preferences of one’s homo economicus. I suggest that the reason why economists intend to give to people what they ‘truly want’ is the result of the third-person perspective they endorse when providing normative assessments. This perspective may however offer a biased diagnosis of the normative issues faced by boundedly rational individuals: rather than focusing on apparent errors of reasoning, I indeed argue that the normative challenge raised by behavioural economics is that our behaviour may be influenced by reasons we are either not aware of (e.g., framing effects) or that we do not accept (e.g., addiction). Behavioural anomalies may therefore matter since our autonomy can be violated: I therefore defend a normative criterion in terms of individual autonomy, according to which it is the ability to choose and accept one’s preferences that matters (chapter 4).

The object of the second part of my thesis is then to provide the basis of a model of preferences that does not rely on a primitive in terms of ‘true preferences’. Rather than considering that the individuals progressively discover some latent coherent preferences, I argue that the
individuals progressively form their own preferences. I show that strategic interactions characterised by a strategic substitutability give an incentive to the players to adopt more aggressive preferences, while games with strategic complementarities tend to generate cooperative preferences. This result highlights the potential impact of public policies on individual preferences: by changing the strategic nature of a game, the government may also change the preferences of the individuals (chapter 5). The general model of preferences I develop in chapter 6 relies on Bacharach’s (2006) variable frame theory and on the theory of team reasoning. I assume that the set of states of the world cannot be unambiguously described, i.e., that there are several equally valid perceptions of the same state of the world. I show that individuals are likely to adopt what Bacharach calls a ‘we-frame’, i.e., to consider themselves as the members of a group and to be actuated by the group objective. Given their initial individualistic perception of the game, players can therefore choose their intentions. Players strategizing with a we-frame are actuated by the collective intention of satisfying the interest of each member of the team, and are therefore team reasoning. I argue that collective intentionality can be formally represented as a choice of collective preferences among team reasoners: I show that team reasoners are likely to choose aggressive preferences with outsiders in games with strategic substitutes, while they tend to adopt cooperative preferences with outsiders in games with strategic complementarities. By identifying themselves as the members of a specific group, the individuals build their own preferences and identity through the choice of the preferences of their group: incoherent and non-selfish preferences are therefore not deviations from some ‘true’ underlying preferences, but the evolving preferences of a socially-embedded agent. I finally show that team reasoning can be interpreted as an ecologically rational heuristic, and more specifically that team reasoning almost systematically outperforms payoff maximising behaviours (chapter 7).

The main conclusion of this thesis is that behavioural findings do not tell us that humans are ‘faulty econs’, and that it falls to the agents—and not to normative economists—to decide whether the incoherence of their preferences matter. Normative economists should instead focus on the process of preference formation, and provide policy recommendations so as to promote individual autonomy, i.e., so as to ensure that the agents have the opportunity to form their own preferences.
REFERENCES


**Guilhem Lecouteux** obtained a PhD in economics from École Polytechnique in 2015, after having graduated from the École Normale Supérieure de Cachan in economics and management. He is currently a teaching assistant at the University of Bristol. His primary research interest is on the problem of how to reconcile normative and behavioural economics, i.e., the normative implications of behavioural findings such as preference incoherence. His research combines game theory, economic methodology, history of economic thought, and philosophy.

Contact e-mail: <guilhem.lecouteux@bristol.ac.uk>
PHD THESIS SUMMARY:
The world as a garden: a philosophical analysis of natural capital in economics

C. TYLER DESROCHES
PhD in philosophy, May 2015
University of British Columbia

This dissertation undertakes a philosophical analysis of “natural capital” and argues that this concept has prompted economists to view nature in a radically novel manner. Formerly, economists referred to nature and natural products as a collection of inert materials to be drawn upon in isolation and then rearranged by human agents to produce commodities. More recently, however, nature is depicted as a collection of active, modifiable, and economically valuable processes, often construed as ecosystems that produce marketable goods and services gratis. Nature consists of various unproduced mechanisms or “natural machines” that are first discovered and then channeled so as to serve human ends. In short, nature as an ideal is a kind of garden that is characterized by natural objects purposefully arranged by intentional human agents.

The first two chapters of this dissertation lays out working definitions of the key terms, such as capital and nature, and then argues that the spatio-temporal particulars denoted by the concept of natural capital, such as ecosystems, are objects (2) capable of producing, (3) depletable, (4) beneficial, (5) original, and (6) self-generative. Among these six characteristics, it is argued that the first four are shared with manufactured capital, while the last two—original and self-generative—drive a wedge between natural and manufactured capital.

Chapter three traces the historical roots of nature as a producer in the works of the Swedish botanist Carl Linnaeus’ (1749) *Oeconomia naturae* and the physiocrats of France during the mid-18th century. This chapter argues that if natural capital is taken to denote nature as an unassisted producer of readily consumable goods and services, then it can hardly be considered a novel concept. There is a nascent category of the concept of natural capital to be unearthed in the writings of classical political economists, such as Adam Smith, John Stuart Mill, and Karl
Marx. When such economic theorists referred to the “spontaneous productions of the Earth” and nature’s “natural products”, they had a distinctive class of production in mind, one that denotes nature’s independently generated products.

Chapter four, entitled, “Critical Natural Capital and Sustainable Development”, tackles a version of the most vexing question concerning natural capital: to what extent can manufactured capital serve as a substitute for natural capital? Economists influenced by the life sciences have long argued that there is a subset of natural capital, critical natural capital, for which there are no substitutes. This special category of natural capital is meant to denote the ecological conditions essential to the continued existence of economic agents and therefore, sustainable development. However, the problem is that no one has explained what these conditions might be and why they are essential for this purpose.

To resolve this issue, this chapter introduces a new theory of what are termed “basic ecological goods” (BEGs). It is shown that BEGs are distinct from ordinary goods in consumer choice theory since the former are objective ecological conditions that must be met for the continued existence of economic agents. BEGs are required for the continued existence of a given agent because they possess objective causal properties essential for this purpose. The upshot of this theory is that the ecological conditions required for human economic activity and, therefore, sustainable development, are no longer shrouded in mystery as they were under the canopy of “critical natural capital”. The theory of BEGs explains what these minimal ecological conditions are and what conditions would have to be met for any good to potentially serve as a substitute for such goods.

Chapter five, “No One Can Preserve Nature”, begins by recognizing a corollary of the garden image of nature. At first glance, this image would seem to entail domesticating every last economically valuable ecosystem to serve human ends. As a result, the status of “wild ecosystems”, “wilderness”, and “untrammeled Nature” are called into question. While this chapter does not consider the desirability or goodness of such a domesticated world, it argues that the preservation paradox is warranted. This paradox, well-known among environmental ethicists, contains three premises: nature is that realm of phenomena that is independent of intentional human agency; preserving and restoring nature requires intentional human agency; therefore, no one can preserve (or restore) nature. While some scholars have argued that
the preservation paradox is misguided, this chapter argues that no one can restore or preserve nature without turning it into an artifact. To defend this claim, this chapter delineates three features that distinguish artifacts from natural objects: artifacts are designed or planned, they possess a function attributed to them by an intentional agent, and they must be modified by an intentional agent. Then, by relying on James Woodward's (2003) analysis of absence causation, it is argued that even those aspects of nature that are merely preserved (where human activity is intentionally omitted) qualify as artifacts.

Finally, Chapter six concludes by acknowledging the limitations of this dissertation and by considering a future direction of research: delineating the moral limits to buying and selling natural capital and ecosystem goods and services. Specifically, an argument is sketched for the moral limits to buying and selling water. Following the recent work of Michael Sandel and Debra Satz on the moral limits to markets, it is argued that, in desperate circumstances, when water is radically scarce, buying and selling water in the marketplace will almost certainly violate what Robert Nozick (1974) refers to as “Locke's Proviso”—a constraint on original acquisitions that requires such activities do not worsen the situation of others.

C. Tyler DesRoches is assistant professor at the School of Sustainability, Arizona State University. He obtained a PhD in philosophy from the University of British Columbia in May 2015, under the supervision of Margaret Schabas. Tyler is a founding editor of the Erasmus Journal for Philosophy and Economics, and has a M.A. in philosophy and economics from the Erasmus Institute for Philosophy and Economics, and a M.A. in economics from the University of Victoria. His articles have appeared in journals such as History of Political Economy, Journal of the History of Economic Thought, and the Canadian Journal of Forest Research.

Contact e-mail: <tyler.desroches@asu.edu>