



ERASMUS JOURNAL FOR PHILOSOPHY AND ECONOMICS
VOLUME 6, ISSUE 1, SPRING 2013

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

François Claveau
C. Tyler DesRoches
Joost W. Hengstmengel
Luis Mireles-Flores
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 Heilmann,
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.

ERASMUS JOURNAL FOR PHILOSOPHY AND ECONOMICS
VOLUME 6, ISSUE 1, SPRING 2013

TABLE OF CONTENTS

ARTICLES

Collective intentionality in economics: making Searle's
theory of institutional facts relevant for game theory
CYRIL HÉDOIN [pp. 1-27]

Early marginalist ideas on money:
some neglected exceptions to the quantity theory
GERMÁN D. FELDMAN [pp. 28-48]

By "fancy or agreement": Locke's theory of money
and the justice of the global monetary system
LUCA JACOPO UBERTI [pp. 49-81]

Hayek on the wisdom of prices: a reassessment
RICHARD BRONK [pp. 82-107]

BOOK REVIEWS

Mary S. Morgan's *The world in the model:
how economists work and think*
ROBERT SUGDEN [pp. 108-114]

Julian Reiss's *Philosophy of economics:
a contemporary introduction*
HAROLD KINCAID [pp. 115-121]

Wilfred Beckerman's *Economics as applied ethics:
value judgements in welfare economics*
JONATHAN B. WIGHT [pp. 122-126]

Andrew M. Yuengert's *Approximating prudence:
Aristotelian practical wisdom and economic models of choice*
RICARDO F. CRESPO [pp. 127-133]

James Halteman and Edd S. Noell's
Reckoning with markets: moral reflection in economics
BETSY JANE CLARY [pp. 134-139]

Ryan Walter's *A critical history of the economy:
on the birth of the national and international economies*
TILL DÜPPE [pp. 140-146]

Samir Okasha and Ken Binmore's *Evolution and rationality:
decisions, co-operation, and strategic behaviour*
WILJAN VAN DEN BERGE [pp. 147-152]

Axel Gosseries and Yannick Vanderborght's
Arguing about justice: essays for Philippe Van Parijs
JULIANA BIDADANURE [pp. 153-160]

PHD THESIS SUMMARIES

Evolutionary foundations of Coasean economics:
transforming new institutional economics
into evolutionary economics
MASAHIRO MIKAMI [pp. 161-164]

Beyond green and mainstream: on the normative
foundations of environmental policy
STIJN NEUTELEERS [pp. 165-168]

Collective intentionality in economics: making Searle's theory of institutional facts relevant for game theory

CYRIL HÉDOIN

REGARDS, University of Reims Champagne-Ardenne

Abstract: Economic theories of team reasoning build on the assumption that agents can sometimes behave according to beliefs or preferences attributed to a group or a team. In this paper, I propose a different framework to introduce collective intentionality into game theory. I build on John Searle's account, which makes collective intentionality constitutive of institutional facts. I show that as soon as one accepts that institutions (conventions, social norms, legal rules) are required to solve indetermination problems in a game, it is necessary to assume a form of collective intentionality that comes from what I call *a common understanding of the situation among the players*. This common understanding embodies the epistemic requirements for an institution to be a correlated equilibrium in a game. As a consequence, I question recent claims made by some economists according to which game-theoretic accounts of institutions do not need to assume collective intentionality.

Keywords: epistemic game theory, common understanding, collective intentionality, John Searle, institutions

JEL Classification: A12, B40, B52, C70

Given that the foundational problem of the multiplicity of equilibria in games has still not been solved, game theorists acknowledge that the extension of Bayesian decision theory to strategic interactions is far from straightforward. While rational choice theory and game theory are still important pieces in the toolkit of most economists, many decision and game theorists are trying to expand the traditional set of assumptions underlying the use of these tools to account for both

AUTHOR'S NOTE: A previous version of this article was presented at the seminar "Institutionnalismes" at the University of Reims (2012) and at the joint conference AHE, IIPPE, FAPE held in Paris (2012). I particularly thank Nicolas Brisset and Fabien Tarrit for their comments. I also thank Herbert Gintis and another referee for their suggestions and remarks. Finally, I thank François Claveau for his helpful editorial advices. Errors and omissions remain mine.

theoretical and empirical issues. The recent introduction of collective intentionality in a game theoretic framework is an illustrative instance of such a tendency.

Theories of collective intentionality have been developed by philosophers for many decades. Their aim is to provide a convincing explanation for the way individuals coordinate and cooperate by pursuing collective goals and by forming collective intentions or beliefs. In economics, theories of *team reasoning* have been also proposed (Bacharach 2006; Sugden 2000; 2003); they build on the assumption that sometimes agents can behave according to beliefs or preferences attributed to a group or a team. The main motivation for the introduction of team reasoning into economics is related to some of the conceptual difficulties faced by game theory, in particular the problem of the multiplicity of Nash equilibria. A player who team-reasons is seeing the game under a different frame; the decision problem he has to solve is not an individual one, but a collective one. This framing process consists for each agent in transforming the level of agency, i.e., to act according to preferences that are not ascribed to an individual (himself) but to a collective (the team). In some cases, team reasoning allows players to coordinate on one determinate equilibrium.

A distinguishing feature of these theories is that they define ‘I-mode’ and ‘we-mode’ of reasoning as substitutes. Though team-reasoning theories are highly innovative and important, I see at least two limits with this last feature: firstly, team-reasoning does not always suppress the problem of the multiplicity of equilibria. For instance, in pure coordination games, using team-reasoning will not help players to coordinate. Secondly, in the case where the interests of players are not perfectly aligned (i.e., in mixed-motives games), it is not clear why players would or should team-reason. More fundamentally, in such cases, players must agree on a common ordering of collective preferences and, until now, very few contributions have tackled this difficult issue.¹

In this article, I propose a different framework to introduce collective intentionality into game theory. I build on John Searle’s theory of institutional facts, which makes collective intentionality constitutive of institutional facts. I show that as soon as one accepts that institutions (conventions, social norms, legal rules) are required to solve indeterminacy problems in a game, it is necessary to assume a form

¹ Two significant exceptions are Sugden 2000; and Hakli, et al. 2010.

of collective intentionality that comes from what I call a *common understanding of the situation among the players*. This concept embodies the epistemic requirements for an institution to be a correlated equilibrium in a game and can be related to Searle's collective intentionality.

This approach is valuable in at least two respects. Firstly, even though team reasoning may explain cooperation and coordination in some circumstances, institutions are arguably a more pervasive and general explanation. Secondly, by introducing collective intentionality in game theory, I connect Searle's theory of institutional facts with economic theory. Despite the fact that Searle's theory tackles important issues regarding the nature of institutions, his approach has generally been either misunderstood or ignored by economists. A recent article by J. P. Smit, Filip Buekens, and Stan du Plessis (2011) illustrates the standard rational choice objection made against Searle's concept of collective intentionality. My article offers some counterarguments to this critique.

The paper is divided into five sections. The first section presents John Searle's account of collective intentionality and notes that it encompasses all cases where individuals behave according to an institution. In the next section, I interpret institutions as correlated equilibria in a game. I make explicit the epistemic requirements needed for institutions to signal to individuals what they ought to do. The third section shows that these epistemic requirements are fulfilled when players have a common understanding of the situation. I suggest that common understanding has the same properties than Searle's collective intentionality. I build on this point to discuss, in the fourth section, the recent critique of Searle's account by Smit, Buekens, and du Plessis (2011) who argue that an economic explanation of institutions does not need a concept of collective intentionality. A final section briefly concludes.

JOHN SEARLE'S ACCOUNT OF COLLECTIVE INTENTIONALITY

Searle has developed his theory of collective intentionality in several writings (see, in particular, Searle 1990; 1995; 2010). This theory is located at the crossroads of Searle's theory of mind and social theory and a full explanation of it would require us to go deeply into the complexities of his accounts of both. Since this is not possible here, I will confine myself to the elements that are the most important when

one attempts to introduce collective intentionality into a game-theoretic framework. In particular, psychological and cognitive features are secondary, while the link between collective intentionality and social facts is of the utmost importance. Still, an essential building block of Searle's account is its naturalistic stance, i.e., that any assumption or explanation regarding linguistic or social phenomena must be consistent with our current knowledge of natural science, in particular physics and biology (Searle 1995, 5-7). As a consequence, a crucial assumption for any acceptable theory of collective intentionality is that any kind of intentional state must be located in individual minds. Moreover, an intentional state (such as a desire, an intention or a belief) must be taken as independent of the actual state of the world; an individual may hold radically wrong beliefs or build his intention on a mistaken conception of the world (Searle 1990). As Searle himself notes, these commonsensical and pre-theoretical requirements are fully in line with all the variants of methodological individualism.

Given these constraints, Searle argues that, in addition to singular or individual intentionality, the ontology of our world is also made of another form of intentionality: collective intentionality. According to Searle, collective intentionality is a *primitive* biological phenomenon that is not an exclusive property of humans but that is also possessed by other animals (Searle 1995, 24). In Searle's social theory, collective intentionality is deemed to be constitutive of any social fact (Searle 1995, 26).

The main implication is that, according to Searle, one cannot reduce collective intentionality to some form of singular intentionality. In particular, Searle (1990) rejects popular accounts of 'we-intentions' that reduce the proposition 'we intend that' to a set of 'I intend that' augmented by a set of mutual beliefs regarding the intentions of everyone, i.e., accounts that equal 'we intend that x' with the infinite hierarchies of propositions 'I intend that x given that I believe that you intend that x, you intend that x given that you believe that I intend that x, I believe that you believe that I intend that x, you believe that I believe that you intend that x', and so on *ad infinitum*, effectively making the intention that x common knowledge in the population.²

² Such a reductive and individualistic account of collective intentionality has been particularly developed by Tuomela and Miller 1988. Note that despite the fact that I use the concept of common knowledge below, my own account is very different from the one of Tuomela and Miller. However, as one of the referees remarked, I shall note that Raimo Tuomela's most recent writings on the topic of collective intentionality

Searle gives several arguments against this kind of reduction. The most decisive one is that while it tries to make collective intentionality the mere aggregation of individual intentional states, it fails to account for the fact that individual intentionality is *derived* from the collective intentionality that members of a group may share. Indeed, in many cases, I intend that *x because* we intend that *x* (or some other propositional content *x'*).

To understand the structure of Searle's framework, it is useful to start with the way Searle defines intentionality. Intentionality is "the capacity of the mind to represent objects and states of affairs in the world other than itself" (Searle 1995, 6-7; see also Searle 2010, 25). According to this definition, intentions, desires, and beliefs are all kinds of intentional states. For example, desires are an instance of a mind-to-world relationship where one wants something to become actually the case. Beliefs are rather an instance of world-to-mind relationship, since a belief is about something that is already the case. In both cases, the intentional state is about an external object.

Though Searle's analysis essentially focuses on a specific form of intentional states, namely intentions-in-action (i.e., intentions that cause an action in a self-referential way), it is perfectly reasonable to extend it to other intentional states. Still, making the simplifying assumption that intentions always lead to the corresponding actions, discussing the case of intentions-in-action is sufficient for our purpose. Searle's analysis of the various forms of intentionality uses the simple notational device *S(p)*, where *S* stands for the psychological or intentional state (an intention-in-action, a belief, a desire) and *p* denotes the propositional content and consequently determines the conditions of satisfaction of the state. In the simple case of an action where one is raising one's arm, we have the following expression, where *ia* stands for *intention-in-action*:

(1)**ia** (this **ia** causes: my arm goes up) CAUSES: MY ARM GOES UP

The first term outside the parentheses denotes the intentional state. The expression inside the parentheses depicts the "mental" component of the action. It states the conditions of satisfaction of the intentional state which, since we are concerned here about an intention, is self-referential: one of the conditions of satisfaction of the state is that the

(e.g., Tuomela 2007) largely give up the individualistic stance and are much closer to the account developed here. Tuomela has also written extensively on how Searle's account and his own are related, e.g., Tuomela 2011.

state itself causes an event of the type represented in the rest of the conditions of satisfaction. The expression in capital letters is a representation of the “physical” component of the action. For the remainder of this section, I will ignore the physical component and assume that it always satisfies the content of the mental component.

The same expression may be used to represent an intention causing an action that causes itself a further action, i.e., a by-means-of relation. It can also represent an intention causing an action which is constitutive of a further action, i.e., a by-way-of relation. As an example of the former, in the case where one intends to fire a gun by pulling the trigger, we have:

(2) **ia** B by means of A (this **ia** causes: A [trigger pulls], causes: B [gun fires])

This reads as ‘I have an intention-in-action to cause B (fire the gun) by means of A (pulling the trigger), this intention-in-action leads me to have the intention to pull the trigger to make the gun fire’.

The point is now to see how this notation and the theoretical framework on which it is grounded can extend to cases of collective intentionality. Here enters Searle’s main contention: while individuals are regularly in a singular-intentional state such as ‘I intend’, ‘I desire’ or ‘I believe’, we also routinely use first-person plural forms of intentionality such as ‘we intend’, ‘we desire’ and ‘we believe’. As noted, Searle takes this form of collective intentionality to be irreducible to singular forms and, moreover, to be *causally* responsible for the formation of individual intentional states (i.e., ‘I intend that... *because* we intend that...’).

Take the example of a play in an American football game where an offensive lineman intends to block defensive players. The lineman’s intention is caused by the fact that he participates in a football game and that it has been decided that the offense he is part of will intend to complete a pass to gain some yards. Originally, Searle (1990) proposed the following general formulation in terms of an achieve-collective-B-by-means-of-singular-A type:

(3) **ia** collective B by means of singular A (this **ia** causes: A [block defensive players], causes B [completion of the pass play])

Here again, the expression outside the parentheses denotes the type of intentional state. Therefore, the collective dimension of the state

features explicitly. It reads as ‘we intend B by means of A’, where B is the collective target of the intention, while A is the individual means. Since for Searle, intentionality is the result of a brain process, this literally means that the collective intentionality is “in the head” of each individual. However, this original account of collective intentionality faces a great difficulty.

Indeed, it should be clear that in the formulation above A can only cause B under certain circumstances, in particular if others have the appropriate intentional states. Therefore, the propositional content inside the parentheses is unacceptable (Bardsley 2006, 137). This difficulty comes from the fact that any account of collective intentionality must satisfy what is generally called in the literature the *uncontroversial constraint*: an individual’s intentions (or any other intentional states) cannot be said to range over others’ actions. As a consequence, it seems impossible to have a collective intention (‘we intend’) or a collective belief (‘we believe’), because one’s intentional state cannot cause others’ actions. Stated in this way, Searle’s account—as well as all other non-reductive accounts of collective intentionality—faces a dilemma: either violating the uncontroversial constraint or giving up the non-reductionist stance (see Bardsley 2006). Both alternatives are hardly desirable, since the former would imply a barely sustainable metaphysical assumption, while the latter would be in contradiction with the essence of a non-reductive account of collective intentionality.

In a full chapter devoted to his theory of collective intentionality, Searle (2010, 42-60) reformulates his account to make clear that the tacit presupposition that everyone will be doing his part in the collective endeavour is not part of the propositional content but rather is an implicit statement constitutive of the collective intentionality. Searle’s new account is thus the following (2010, 53):

- (4) **Bel** (my partners in the collective also have intentions-in-action of the form (**ia** collective B by means of singular A (this **ia** causes: A [block the defensive players], causes: B [completion of the pass play])))

We have thus added an extra clause **Bel** that reads as follows: ‘I have a belief that my partners in the collective also have the same collective intention-in-action as mine’. Searle underlines that this belief is not a part of the *content* of the intention-in-action, thus satisfying the uncontroversial constraint. Rather, it is an integral part of the collective

intention: to have a collective intention implies to believe that others have the same collective intention. Incidentally, reading Searle literally, this belief seems *individual*: it is about others' intentional states but it is not a 'we believe that' type.

I do not regard this new account as fully satisfactory. First, we-intentionality can no longer be considered non-reducible because individual beliefs are necessary. Second, and more importantly, the extra clause **Bel** seems insufficient: even if I believe that you have the same collective intention than me, I may doubt that you have the same belief about me. In this case, I may refrain to form an individual intention on the basis of this collective intention. By this reasoning, we reach the inescapable conclusion that for a collective intention to be effective, we must have an infinite set of iterated mutual beliefs about our intention, ultimately grounding the collective intention on a singular form of common belief. As I suggest in the next two sections, there is one escape though if we interpret Searle's **Bel** condition not in the traditional epistemic sense but as a dispositional and pre-intentional property. In this case, the very possibility of collective intentionality is a feature of what Searle calls the "Background". I further discuss this point below.³

Crucially, collective intentionality (in particular collective intentions) is constitutive of institutional facts.⁴ According to Searle's taxonomy of facts (Searle 1995, 121), social facts involve a form of cooperation which requires collective intentionality. Since institutional facts are a subclass of social facts where status functions are assigned to brute phenomena (i.e., 'this piece of paper counts as a twenty dollar bill'), they also rely on a non-reductive form of collective intentionality. In particular, Searle argues that the maintenance of institutions through time necessitates

³ Searle's new account also features a more explicit discussion of what he calls "collective recognition" or "collective acceptance". He defines collective recognition as a weaker form of collective intentionality which operates even when social agents do not have intentions to cooperate. Collective recognition is an important category because it underlies all of Searle's theory of institutional facts, in particular his concepts of status functions and of constitutive rules. Unfortunately, this contributes to muddle the water somewhat, in part because it fits uneasily with Searle's own framework developed in *The construction of social reality* (1995). In this book, Searle explicitly indicates that cooperation is constitutive of all social facts, of which institutional facts are a subset. Since Searle's definition of cooperation is fairly large (see Searle 1995, 23-24), it is difficult to point to (non-social) facts involving collective recognition but not collective intentionality.

⁴ Note that Searle's theory of institutional facts also makes the strong claim that language is also constitutive of institutional facts. For a function to be assigned, it has to be represented by some form or another of symbolism. I will not put this idea under scrutiny here but I will have to go back to it in the last section.

collective intentionality (Searle 2010, 57). In the next two sections, I will show that a similar non-reducible account of collective intentionality is also relevant in a rational choice framework.

INSTITUTIONS AS CORRELATED EQUILIBRIA: MUTUAL AWARENESS AND SYMMETRIC REASONING

Building on Searle's account of collective intentionality, this section and the next demonstrate that it is fruitful to assume collective intentionality in rational-choice and game-theoretic frameworks. In fact, it is even necessary if one accepts the necessity of institutions to account for coordination and more generally for agents' choices in strategic interactions.

Our socioeconomic life is full of situations mixing coordination and conflict. Even if pure coordination problems are also pervasive, social interactions where agents have to coordinate while having conflicting interests are particularly difficult and important. The difficulty is that in such interactions, efficiency makes coordination desirable but equity concerns and conflicting preferences over outcomes may make coordination particularly hard to achieve.

Social scientists, including economists, tend to agree that *institutions* are one of the main devices used by agents in mixed-motive games to form beliefs and to act. Institutions are usefully defined as *social norms, conventions, legal rules or formal organizations that generate a set of consistent beliefs in a population such that a stable behavioural pattern is observable through time*. In other words, institutions are kinds of social objects that enter as an input in the agents' intentional states to help them form consistent beliefs regarding the state of the world and regarding what others will do.

In the remaining of the paper, I will follow Herbert Gintis (2009) in formalizing institutions as *correlated equilibria* in games. The main motivation for using this solution concept rather than the more classical Nash equilibrium solution is twofold. Firstly, as Aumann and Brandenburger (1995) have emphasized, the epistemic conditions that must be satisfied for rational players to play a Nash equilibrium are very demanding, especially with three or more players. Indeed, both common knowledge of expectations and common priors over the way the game will be played must be satisfied. Except under highly specific circumstances, these conditions are unlikely to be satisfied. Secondly, and more fundamentally, norms and other kinds of

institutions are more properly interpreted as implementing correlated equilibria. An important theorem by Robert Aumann (1987) shows that the strategy profile played by Bayesian rational players with common priors forms a correlated equilibrium. If one accepts the assumption that common priors are a necessary condition for an institution to exist, then formally institutions are correlated equilibria.⁵

My main claim is that for an institution to serve as a correlating device in a game, we have to assume a form of collective intentionality qualitatively similar to Searle's. I will work through the following mixed-motive game, which I will call the 'property game' (see Table 1).

Table 1: The property game

		Bob	
		attack	negotiate
Ana	attack	$(V-C)/2; (V-C)/2$	$V; 0$
	negotiate	$0; V$	$V/2; V/2$

The property game (also known as the chicken game) depicts an interaction where two agents are fighting for an asset (a prize, a territory) of value V . Each player has two strategies available: either to attack in order to take the asset by force or to negotiate peacefully. When one of the players attacks while the other attempts to negotiate, the former wins the prize for sure. If both negotiate, they share the asset equally. Finally, if both attack, each player wins with probability $\frac{1}{2}$ and the loser suffer a loss of $-C$ which might be interpreted as an injury. We assume that $C > V$.

As usual, this game structure is common knowledge among the players, the players are rational and this is common knowledge. This game embodies an element of coordination that is captured by the fact that it has two Nash equilibria in pure strategies: [attack; negotiate] and [negotiate; attack]. There is also a third, mixed-strategy Nash equilibrium where each player attacks with probability V/C . Because $C > V$, each player prefers losing the asset rather than fighting when the

⁵ Note that using the concept of correlated equilibrium does not entail the rejection of the Nash equilibrium solution, nor does it imply making non-standard assumptions regarding rationality or knowledge in a game. Indeed, every correlated equilibrium corresponds to a Nash equilibrium in the larger game where Nature moves first to signal the state of the world to the players.

other attacks. However, the players' preferences over the outcomes are clearly inconsistent since each prefers to take the asset over giving it up.

The difficulty lies in the fact that it is common knowledge among the players that their preferences over the outcomes are inconsistent. In such a case, an institution (a social norm or a legal rule) is required to settle the dispute. Typically, an institution will select one of the equilibria; however, in our case, there is no reason why an institution should adjudicate the conflict between Ana and Bob in favour of the former or of the latter. Indeed, who will be *entitled* to claim the asset depends on the specific situation of the two players. For instance, the norm might distinguish between the *actual* possessor of the asset (the incumbent) and the non-possessor (the challenger). Assume that the distinction between incumbent and challenger is non ambiguous and common knowledge and assume that each player is equally likely to be the incumbent or the challenger; we can then provide the alternative description of the game represented in Table 2.

Table 2: The property game with incumbent and challenger

		Bob			
		always attack	always negotiate	attack if incumbent	attack if challenger
Ana	always attack	$(V-C)/2$	V	$(3V-C)/4$	$(3V-C)/4$
	always negotiate	0	$V/2$	$V/4$	$V/4$
	attack if incumbent	$(V-C)/4$	$3V/4$	$V/2$	$(V-C)/2$
	attack if challenger	$(V-C)/4$	$3V/4$	$(V-C)/2$	$V/2$
N.B. Only the payoffs of the row player are shown.					

Once the players are able to distinguish between being the incumbent and being the challenger, they can use *conditional* strategies, i.e., strategies that associate an action to a state of the world. Framing the 'property game' this way does not solve the problem of the multiplicity of equilibria since there are still two equilibria in pure strategy⁶ and one equilibrium in mixed-strategy. However, now that

⁶ These are: [attack if incumbent; attack if incumbent] and [attack if challenger; attack if challenger].

the distinction between incumbent and challenger has been made, an institution that settles the dispute along an intuitively more satisfactory impersonal criterion can be defined. In particular, a *property norm* will select the [attack if incumbent; attack if incumbent] equilibrium, avoiding costly conflicts that arise with probability $(V/C)^2$ when the players play the mixed-strategy equilibrium.⁷

I will now examine how the property norm effectively changes the players' beliefs. To acknowledge the role played by the features regarding the knowledge and the beliefs of the players, it is useful here to define the *epistemic* property game. An epistemic game retains the classical feature of a game (i.e., a set of players, a set of pure strategies and a consequence function), but adds an explicit mapping of the players' knowledge about the states of the world (see Gintis 2009, 83-84).

Define Ω as the set of possible states of the world and P_i the possibility sets of Ω for each player i . Possibility sets partition the set of worlds Ω into units of knowledge defined by a knowledge partition. A possibility set indicates which states of the world $\omega' \in \Omega$ a player thinks is possible when the actual state of the world is $\omega \in \Omega$, which we denote as $P_i\omega$. Finally, each player is endowed with a subjective prior $p_i(\cdot; \omega)$ that is a function of the state of the world ω . A player's subjective prior defines, among other things, his belief over the strategy profiles to be played by the other players at ω . A Bayesian rational player will choose the strategy that maximizes his expected payoff given his conjectures regarding the state of the world and how others will play. Finally, we identify an *event* E as the set of possible worlds in which that event happens. An agent *knows* an event E when every world the agent considers possible is a subset of E . The event of an agent i knowing that E is denoted K_iE .⁸ Formally, $K_iE = \{\omega \mid P_i\omega \subseteq E\}$.

Now apply the above formalism to the 'property game'. We distinguish between two states of the world $\Omega = \{\omega_1; \omega_2\}$ which obtain with an equal probability $q_\omega = 1/2$. Assume that ω_1 is the state of the

⁷ There is an evolutionary argument for the selection of one of the equilibria with conditional strategies beyond the mere intuitive plausibility. Indeed, in an evolutionary setting, the mixed-strategy is not evolutionary stable. Quite the contrary, both conditional strategies are evolutionary stable. An intriguing fact however is that the anti-property strategy 'attack if challenger' has an even chance to spread in the population than the more intuitive property strategy.

⁸ K is a knowledge operator. As usual in epistemic game theory, I assume that the knowledge operator satisfies the requirements of the modal logic S-5. See Binmore 2007, chapter 12, for an accessible presentation.

world where Ana is the incumbent while she is the challenger in ω_2 . My assumption that the distinction incumbent/challenger is unambiguous is equivalent to the formal statement that ω_1 and ω_2 are in different cells in the knowledge partition of both players, i.e., when the actual state is ω , each player knows this for sure. Therefore, the events $E = \{\omega_1\}$ and $E = \{\omega_2\}$ are *public events*. When E happens, it is self-evident for each player that E happens, i.e., $E = K_i E$ for all i . Since ω_1 and ω_2 are also in two different cells in the communal possibility sets of all i ,⁹ E is common knowledge because everyone knows that E is self-evident for everyone: i knows that E obtains, i knows j knows that E obtains, i knows that j knows that i knows that E obtains, and so on. Formally, $E \subseteq CK(E)$; for all E and with $CK(E)$ the event that (Everyone knows that) $^\infty$ E . The property norm described above may now be restated as follows:

$$N(E) = (\text{if } \omega_1, [\text{attack}; \text{negotiate}], \text{if } \omega_2, [\text{negotiate}; \text{attack}])$$

Our notation $N(E)$ indicates that event E is governed by norm N (see Gintis 2009). Generally speaking, a norm $N(E) = (s_1, \dots, s_n)$ specifies a vector of strategies for the n players such that no Bayesian rational player can gain by playing another strategy s_i' when E obtains. The norm defines a *correlated equilibrium* because the strategies are correlated to a specific state of the world. Crucially, at equilibrium, the norm $N(E)$ itself is common knowledge among the players, thus $CK(E) \subseteq CK(N)$.

This expression depicts an *indication relation*. To say that an event E is governed by a norm N is the same as to say that E *indicates* N to the players. The notion of indication relation is central in David Lewis's theory of common knowledge (Lewis 2002 [1969]; Cubitt and Sugden 2003) and captures the fact that individuals have to be able to infer a norm from a particular event for that norm to be effective.

This framework makes explicit the epistemic properties of an institution. The players' subjective priors lead them to form a conjecture ϕ_i^ω regarding how the other players will play at ω . Thus, we can define a belief profile $\phi(\omega) = (\phi_1^\omega, \dots, \phi_n^\omega)$ correlated with the state ω , such that Bayesian rational players will respond by a strategy profile $s(\omega) = (s_1^\omega, \dots, s_n^\omega)$ where everyone is maximizing expected utility. In the 'property game', the property norm leads everyone to conjecture

⁹ The communal possibility sets are defined as the *meet* of the possibility sets of each player. Obviously, since the possibility sets of all players distinguish ω_1 from ω_2 , the communal possibility sets have also this property.

that the incumbent will attack: it correlates the state of the world to a set of conjectures and of strategies, i.e., $N(E) = \phi(\omega)$. Thus, we have a formal expression of the ability of an institution to generate a set of beliefs in a population.

However, this is not the end of the story. We have assumed the indication relation $CK(E) \subseteq CK(N)$ while in fact it embodies crucial epistemic assumptions. Specifically, at least two additional assumptions that go beyond the classical framework of Bayesian decision theory are needed (Gintis 2009): *mutual awareness* and *symmetric reasoning*. I argue that both can be accounted for as properties of a form of collective intentionality such as envisioned by Searle.

The first thing to note is that, in my example, all events are public events. That means that when E obtains, each player knows for sure that E obtains. But to be allowed to infer from the mutual knowledge of E that E is common knowledge, we have to make the seemingly benign assumption that each player knows the knowledge partition of the other player. This assumption is implicit in the above model because a description of each state of the world includes, among other things, a description of the agents' possibility sets. However, this mutual awareness assumption may be really strong in some settings. In my example, it consists in hypothesizing that the incumbent/challenger distinction is *really* unambiguous and that Ana and Bob both know that they exactly define in the same way both concepts. Moreover, it means that if Ana knows she is the incumbent, she must be able to infer with a total confidence that it is unambiguous for Bob that he is the challenger.

Regarding symmetric reasoning, it is crucial to note that the ability for a norm $N(E)$ to become common knowledge when an event E is publicly observed relies on the fact that everyone infers N from E . We say that two persons i and j are symmetric reasoners if and only if i can infer from his knowledge that x and that j knows that x , that j also knows y . More formally, for all i, j and for any events $E, N \subseteq \Omega$,

$$(K_i E \subseteq K_i N) \wedge (K_i [K_j E]) \Rightarrow K_i [K_j N].$$

Such an assumption is at the heart of the indication relation pioneered by Lewis, who supposes that individuals share a set of modes of reasoning (Cubitt and Sugden 2003; Lewis 2002 [1969]; see also Gintis 2009 and Vanderschraaf 1998). Moreover, considering that individuals can deliberate over the modes of reasoning used by members of a

population, we have a second-order problem of coordination where each person is uncertain regarding the way others reason. The problem of course is that there is no reason to suppose that there is only one way to infer y from x , even if deduction is available. People routinely use other modes of inference as a function of the situation in which they interact.

One way to circumvent the problem of mutual awareness and symmetric reasoning would be to assume that the players in an epistemic game have *common priors*. The common prior assumption (also known as the Harsanyi doctrine) simply consists in taking for granted that all players i in a population share the same subjective prior $p_i(\cdot; \omega)$. It is then a theorem that Bayesian rational players with common priors will implement a correlated equilibrium in a game where Nature chooses the state of the world (Aumann 1987). As noted above, this theorem is the main motivation for using the concept of correlated equilibrium. But the common prior assumption is clearly disputable, at least when we do not restrict ourselves to phenomena where objective probabilities can be inferred from law-like regularities.¹⁰

In particular, combined with the assumption that individuals' posterior beliefs are common knowledge, the Harsanyi doctrine leads to the claim that it is impossible to “agree to disagree” (Aumann 1976). This claim clearly does not stand up against empirical evidence and must be considered more a formal curiosity than a substantive insight. Still, it remains true that people are largely able to coordinate in their daily activities and even on more fundamental problems such as agreeing on principles of justice (e.g., Binmore 1998). What I want to suggest is that something akin to the Searlian collective intentionality might provide a satisfactory explanation of mutual awareness and symmetric reasoning.

COMMON UNDERSTANDING AND COLLECTIVE INTENTIONALITY IN GAMES

Since David Lewis's study of conventions (Lewis 2002 [1969]), it is standard to assume that norms and more generally institutions are devices that select a Nash equilibrium in a game. However, I follow Gintis (2009) in considering that it may be more relevant to define

¹⁰ This is the essence of Binmore's (2009) critique of Bayesianism. Following the seminal work of Leonard Savage, Binmore contends that Bayesian decision theory only operates on secure grounds when applied to small worlds, *i.e.* worlds where we know that we cannot be surprised. In large worlds, assuming common priors and asserting the absolute validity of Bayesian updating of subjective probabilities is foolish.

institutions as correlated equilibria. My main motivation in doing so is methodological: to define an institution as a correlated equilibrium leads one to reflect on the epistemic properties of the institution. As the preceding section showed, this approach reveals non trivial epistemic assumptions routinely made in game theory, in particular mutual awareness and symmetric reasoning. I now suggest that one can interpret mutual awareness and symmetric reasoning as peculiar instances of Searle's collective intentionality.

Hédoin (2012) proposes to frame the assumption of symmetric reasoning as referring to cases where agents have a *common understanding of the situation*. Basically, the point is that to infer a set of conclusions C from a situation defined by an event E , the members of the population will use a set of modes of reasoning and of background information B . If every member of a population infers C from E on the basis of B , then we say that members of the population have a *shared-reflexivity* of the situation. More formally,

(Φ - P)-shared reflexivity: there is (Φ - P)-shared reflexivity among the members of a population constituted by the set of agents P if the agents share the same reasoning modes and background information B with respect to a non-empty set of events Φ . In practice, all agents in N infer the same conclusion C when they observe an event $E \subseteq \Phi$.

If for each person C contains propositions about the conclusions reached by the other members of the population, then C depends on the set B_i used by each member i of the population. We define a *second-order shared-reflexivity of the situation* in the population as the fact that each person j embeds in B_j information regarding the sets B_i of every other person i . In principle, this leads to an infinite degree of order of shared reflexivity. There is *common understanding of a situation* among the members of a population when the degree of order of shared reflexivity reaches an arbitrarily high level. Formally,

(Φ - P)-common understanding: there is (Φ - P)-common understanding among the members of a population constituted by the set of agents P if there is an infinite order of (Φ - P)-shared reflexivity. In practice, all agents in P infer the same conclusion C when they observe an event $E \subseteq \Phi$ and this is common knowledge.

When common understanding is obtained in a population, members of this population frame the situation identically and crucially take this fact for granted. Common understanding is a necessary condition for something to become common knowledge in a population. Common understanding is a cognitive disposition that allows the formation of intentional states such as beliefs or intentions. In this sense, it is partly *pre-intentional*. I suggest that the concept of common understanding can be extended to mutual awareness. Indeed, mutual awareness corresponds to proposition: $K_i E \subseteq K_i [K_j E]$.

This proposition says that if i knows E , then he also knows that j knows E . The proposition is grounded on an inference that what is self-evident to i is also self-evident to j . Therefore, mutual awareness is no more than a special case of symmetric reasoning defined by an indication relation linking two separated events.

Where common understanding comes from is surely an important question. It is quite probable that it is partly innate. As members of the human species, we share a common evolutionary history and consequently basic cognitive functions and mechanisms. As an example, humans are universally disposed to identify recurrent patterns in nature. However, our propensity to interpret in similar ways public events (and our propensity to be mutually aware of such events) surely rely on the fact that the members of a population will generally share a system of symbolic cues and meaning (Chwe 2003). Hence, common understanding is largely cultural. In particular, common understanding will generally obtain among the members of the same *community*, i.e., a group of persons who interact through the same institutions. By interacting through the same institutions, individuals progressively develop cognitive and behavioural dispositions constitutive of their modes of reasoning.

The apparent circularity of my argument, where common understanding allows the working of institutions and institutions foster common understanding is not a problem. Indeed, the pre-intentional content of common understanding refers to what Searle calls the “Background”, which he defines as “the set of non-intentional or pre-intentional capacities that enable intentional states of function” and where capacities are “abilities, dispositions, tendencies, and *causal structures generally*” (Searle 1995, 129, emphasis in the original).

Searle argues that in the first place the Background enables linguistic interpretation: the set of truths conditions regarding the meaning of

a sentence is always determined given certain Background capacities. But the pre-intentional structure of the Background is also essential in enabling what I propose to call “institutional interpretation”: the social world is full of institutional signs (conventions, rules, norms, signals) that must be interpreted. In the ‘property game’, the content of the property norm has to be interpreted by Ana and Bob. As we have seen, game theory makes this point obscure with the hidden assumptions of mutual awareness and symmetric reasoning. The point is that an institutional sign is always incomplete and must be *interpreted*. Interpretation always takes place on the basis of a more global set of institutions that defines a community. Searle’s Background is captured by the pre-intentional content of the common understanding concept.

Undoubtedly, common understanding also has a conscious and intentional content. To illustrate this claim (which could be easily tested experimentally), a simple appeal to one’s occasional experience of interacting with people endowed with a different culture should be sufficient. When I drive from my home to my University office, I generally follow the institutional signs on the road (i.e., the traffic rules) without much deliberation. I believe that other drivers will stop at red lights and drive on the right because I *take it for granted* that they are acquainted with such institutional signs. In other words, because we are members of my community, I do not have to reflect profoundly on my interpretation of the traffic rules. Now, things are clearly different when I drive my car in other countries. Here, even though institutional signs might be in principle identical (such as red lights), I will (consciously) be much more thoughtful in their interpretation. Basically, since I *know* that I interact with “strangers”, I cannot fully assume common understanding and I will form beliefs regarding their interpretation of institutional signs. Here, I want to suggest that this type of intentional content that is constitutive of common understanding is the product of a collective intentionality of the type conceptualized by Searle.

When a common understanding of a situation is obtained, each participant in the interaction takes for granted that everyone else is interpreting the situation like him.¹¹ This pre-intentional commitment

¹¹ It is possible to express the idea of common understanding in purely individualistic terms, for example by assuming that symmetric reasoning is common knowledge. In the standard epistemic game-theoretic approach used in the preceding section, each possible state of the world contains everything that is relevant for the players, including possibly the kinds of reasoning. Therefore, once we have partitioned Ω into

sets the ground for a *collective* intentional state which can be formulated with the following proposition:

- (5) *Collective intentionality in a game*: as members of population P , in the situation defined by the game G and a specific event E , we intend to play the strategy profile s^* , therefore I should intend to play the strategy s_i^* .

This statement combines two kinds of reasoning: a we-reasoning where each person, as a member of P , ascribes a collective intentional state to the group of participants in G ; and an I-reasoning where each person reflects on what he should do given his personal preferences and his collective intentional state. The locution ‘therefore’ in this statement embodies a causal relationship where one does s_i^* , because everyone is collectively intending s^* when E holds in G . In the ‘property game’, observing event E , Ana will do s_i^* (attack or negotiate) *because* it maximizes her expected utility given the fact that she takes for granted that Bob and herself collectively intend $N(E)$ in E .

This is in line with Searle’s account of collective intentionality where collective intentional states are causally responsible for individual intentional states. We can make this clearer by reformulating Ana’s and Bob’s intentions to attack or to negotiate in Searle’s notation. The ‘property game’ embodies the kind of general cooperation that is constitutive of social facts according to Searle. What we have here is the following:¹²

- (6) **ia** collective B by way of A (this **ia** causes: A [playing my part of the strategy profile defined by the property norm], constitutes B [cooperation through the property norm])

Since the *content* of the property norm is in the propositional content of (6), both players take it as given. Now, if we add the **Bel** clause, we obtain:

several states of the world and defined the players’ information partitions, all we need to solve the game is already there. As we have already indicated, the same result occurs if we assume common priors. The problem is that there is no reason to assume that the set Ω can be partitioned into well-defined possible worlds or that, at each possible world ω , everyone knows everyone’s possibility sets or modes of reasoning *prior* to any form of interaction or collective recognition.

¹² Since, properly speaking, a set of individual strategies is constitutive of a strategy profile and hence of an institution, the type of collective intention relevant here is not a causal ‘B by means of A’, but rather a constitutive ‘B by way of A’. Searle’s framework allows for the two formulations. This makes no real difference here.

- (7) **Bel** (my partners in the collective also have intentions-in-action of the form (**ia** collective B by way of A (this **ia** causes: A [playing my part on the strategy profile defined by the property norm], constitutes B [cooperation through the property norm])))

Following what has been said in the first section, we must take this belief to be grounded on a pre-intentional disposition. Because the content of this belief includes the collective intention-in-action (6) and since (7) contains the content of the norm, the players actually implement the norm *because* they take for granted that everyone shares the same intention to implement this particular norm with this particular content. Propositions (5) and (7) are two different statements of the same collective intention.

In comparison to theories of team reasoning (Bacharach 2006; Sugden 2000), my account does not need to stipulate team preferences. Indeed, *there is* collective intentionality in the ‘property game’ despite the fact that the matrix only indicates individual payoffs related to individual preferences. Hence, my account avoids one of the most contentious point of the theory of team reasoning, i.e., how team preferences are generated. By contrast, collective intentionality in games helps to explain how the players form their beliefs.

As indicated in the preceding section, an institution $N(E)$ is formally equivalent to a vector of conjectures correlated to the state of the world $\phi(\omega) = (\phi_1^\omega, \dots, \phi_n^\omega)$. As an example again, in the ‘property game’, Ana believes with probability one that Bob will play ‘negotiate’ when ω_1 ; for the same state of the world, Bob believes with probability one that Ana will play ‘attack’. These *individual* beliefs are generated by the fact that both players have a common understanding of the game and that they share a *collective* intentional state in the situation such that a peculiar property norm applies. Collective intentionality in games is thus a plausible explanation of subjective prior, and in particular of *common* prior.

To end this section, note again that the **Bel** clause in proposition (7) is not a belief in the traditional epistemic sense. Therefore, consistent individual beliefs are formed as a result of the common understanding of the situation which is more likely to obtain when people are ostensibly members of the same community. The pitfall of infinite regress is avoided as soon as one acknowledges that common understanding is ultimately grounded on pre-intentional states.

Incidentally, Searle's earlier writings on collective intentionality linked the ability to form collective intentional states with Background capacities: "it seems to me that the capacity to engage in collective behaviour requires something like a pre-intentional sense of "the other" as an actual or potential agent like oneself in cooperative activities" (Searle 1990, 19). Undoubtedly, it is tempting to postulate a deep causal relation between the 'pre-intentional sense of the other' and the peculiarity of the evolutionary history of the human species which has been shaped by the importance of communities (see Bowles and Gintis 2011). Communities are thus at the same time ultimate and proximate causes in our ability to coordinate and to cooperate.

COLLECTIVE INTENTIONALITY AND INSTITUTIONS: SEARLE VERSUS ECONOMIC THEORY?

In this last section, I will make use of what I have said above to discuss the recent critique made by J. P. Smit, Filip Buekens, and Stan du Plessis (2011) against Searle's theory of institutional facts. Their basic claim is that all the Searlian apparatus, including his account of collective intentionality, is unnecessary to explain and understand institutional objects. As an alternative, the authors propose an approach that is deemed to be more compatible with standard economic theory, and which is grounded on the concepts of actions and incentives. I will not review here the various arguments made by these authors against Searle's theory.¹³ I will focus on the authors' alternative approach and claim that while interesting it is not a true alternative to a theory of institutional facts embodying an account of collective intentionality.

The central tenet of the authors' "economic" theory of institutional facts is that institutional facts are tied to a set of actions. Actions are themselves triggered by a set of incentives, i.e., a measure of the relative desirability of the components of a set of possible actions. The authors insist that the origins of the incentives do not matter to characterize the

¹³ Smit, Buekens, and du Plessis's main critique is precisely directed against Searle's account of collective intentionality. They argue that because Searle gives a role to collective intentionality in the assignment of functions, institutional facts cannot be reduced to natural facts. The authors deny this irreducibility and contend that "the mind-dependence of such objects [institutions and institutional facts] is the result of the need for incentivization, and the fact that talk about incentivization is warranted only if talk about desires and beliefs—and hence minds—is warranted" (Smit, et al. 2011, 4). Note that the authors' claim that Searle's theory is ultimately non-reductionist runs contrary to Searle's claim that his theory is grounded on a naturalistic stance.

nature of the institutional objects. In other words, the source of the incentive may be tied to different causal mechanisms for different persons; what counts is that as a result of the incentive everyone is behaving in a particular way that defines the institutional object. To illustrate their point, the authors start from one of Searle's favourite examples, the emergence and the maintenance of a border. I will myself use Smit and co-authors' treatment of this case to make my own point. Here is Searle's discussion of borders as institutional facts:

Consider for example a primitive tribe that initially builds a wall around its territory. The wall is an instance of a function imposed in virtue of sheer physics: the wall, we will suppose, is big enough to keep intruders out and the members of the tribe in. But suppose the wall gradually evolves from being a physical barrier to being a symbolic barrier. Imagine that the wall gradually decays so that the only thing left is a line of stones. But imagine that the inhabitants and their neighbours continue to *recognize* the line of stones as marking the boundary of the territory in such a way that it affects their behaviour. [...] The line of stones performs the same *function* as a physical barrier but it does not do so in virtue of its physical construction, but because it has been collectively assigned a new *status*, the status of a boundary marker (Searle 1995, 39-40; emphasis in the original).

Smit, et al. (2011, 8-9) offer the following alternative account of borders as institutional objects. They ask us to imagine two individuals who I will name, for the sake of consistency, Ana and Bob. These two individuals are lost on a desert island. After a dispute, Ana warns Bob that if he enters this half of the island she will beat him up. To give some substance to her words, Ana simultaneously points to two rocks on opposite sides of the island, cutting the island in half. Bob angrily replies to Ana that the same will be true for her if she enters his half of the island. Assume that the threats are credible. Given this assumption and provided that incentives to ignore the credible threats are insufficient for both individuals, then the invisible line drawn by the two rocks will actually function as a border. According to the authors:

It would be difficult to deny that a border had been set up on the island. But this has happened in a way that violates the essence of Searle's view. Firstly, the requirement for collective intentionality has not been met, as *all the relevant thoughts and claims can be expressed using the singular 'I'*. Secondly, nowhere is reference needed to any irreducibly social facts, objects, or properties. Both

actors can understand the situation fully by using concepts like ‘line’, ‘crossing’, ‘probability of getting beaten up’, etc. [...] We need nothing beyond an understanding of the incentives and beliefs of the two actors in order to grasp the situation fully (Smit, et al. 2011, 9, emphasis in the original).

The authors’ account makes the creation of the border depend on the effectiveness of incentives. Whatever the source of the incentives, as soon as they actually deter Ana and Bob to cross the imaginary line, a border has been created. In this case, contrary to what Searle argues, collective intentionality is not constitutive of the creation of the border.

I think that the authors’ argument, though perfectly in line with a standard rational choice or game-theoretic account, is partially misguided. It fails both to properly state Searle’s approach and to acknowledge the hidden epistemic requirements for their explanation to be valid. I will not delve into the details of the first point. It is sufficient to note that the authors appear to underestimate the role of language in Searle’s theory of institutional facts. Searle has forcefully argued that there cannot be institutional facts and institutions without language, i.e., that language is constitutive of the institutional reality.¹⁴ A language is a set of symbolic devices that represent something beyond themselves. A language helps to convey meaning because it consists of symbolic devices that are *publicly* understandable.

Contrary to what Smit, et al. seem to presuppose, the institutional object (or fact) ‘border’ is tied to the existence of a linguistic device because the individuals must be able to use symbols to be able to form an intentional state with the border concept as a part of the propositional content. Or, in other words, individuals need to be able to represent what a border is before being able to desire or not to cross the border. Moreover, to define the concept of border, Ana and Bob will need other concepts such as ‘line’, ‘rock’, ‘to cross something’, and so on. To use Searle’s terminology, a border is a ‘language-dependent fact’ because it rests on a ‘language-dependent thought’, a thought that it is impossible to have without a language. Obviously, Ana and Bob do not need a language in Smit, et al.’s story; the authors *assume* that the border appears as soon as Ana and Bob, for whatever reason, do not cross an invisible line. The imaginary conversation in their story

¹⁴ See, in particular, Searle 1995, 59-78.

is purely rhetorical, Ana and Bob could be language-less animals and it would still be appropriate to argue that a border has emerged.¹⁵

This leads to my second point. What the authors propose as an alternative account for institutional facts is a standard game-theoretic explanation. What I have said in the preceding two sections must suggest that they are wrong to conclude that their account is exempt from assuming any kind of collective intentionality. The ‘property game’ is easily modified to fit the border story. As soon as Ana has uttered the sentence ‘If you enter this half of the island I will beat you up’ while drawing a fictive line and Bob has answered similarly, the players have fully designed a set of possible states of the world Ω where each state ω is defined by the relative positions of Ana and Bob on the island.

Ana and Bob sentences are examples of performative illocutionary acts which Searle calls *declarations*: the propositional content of the acts (the semantic content of the sentences) is brought into existence by the very performance of the act. By uttering their (credible) threats through a linguistic device, Ana and Bob are doing three things at the same time: they partition Ω into several states ω ; they design their respective possibility sets P_i and their associated knowledge partitions; and they define an institution-as-correlated-equilibrium $N(E) = \phi(\omega)$. As I have argued above, all of this is common knowledge as soon as Ana and Bob have a common understanding of the situation, thus allowing us to assume mutual accessibility and symmetric reasoning. Obviously, this common understanding is permitted because Ana and Bob form a community constituted by the minimal fact that they share an institution, i.e., a language. If my argument that common understanding is analyzable as a form of collective intentionality is correct, then Smit and his co-authors are wrong when they claim that their account is free from any form of collective intentionality.

A possible answer to this argument is that, since language is unnecessary in the authors account, we can still assume that Ana and Bob progressively *learn* to not cross the line, with a border emerging as a result. In this type of evolutionary explanation, we do not need the epistemic apparatus beyond really basic assumptions

¹⁵ Note that I am not arguing that Smit, et al. are necessarily wrong to define an institution as something that may emerge or exist without a language. Nonetheless, first, note that this is not how Searle defines an institution. Second, if we assume that collective intentionality depends on the existence of a language, then to define institutions as language-independent objects is clearly problematic. This last point is, however, an empirical issue.

regarding the learning mechanism (such as reinforcement learning, for example) because agents are assumed to be cognitively unsophisticated. This type of evolutionary framework works well to account for the evolution of animal behaviour.

In my opinion, ‘spontaneous order’ types of explanation are perfectly reasonable, in particular to account for the emergence of institutions. But they complement, rather than compete with, *eductive* explanations such as the one I have developed here. Eductive explanations are more relevant to account for the artificial construction of some low scale institutions (such as a business firm) and also to understand how widely acknowledged institutions are maintained through times.

CONCLUDING REMARKS

In this article, I have presented a framework for introducing collective intentionality into game theory. Though I sympathize with the recent theories of team reasoning arguing for the possibility of team preferences, I think it is rather at the level of the formation of beliefs and intentions that collective intentionality plays a significant role. Basically, what I have been arguing is that when players have a common understanding of the situation, then they have common priors. Institutions can then be defined as correlated equilibria on which players can coordinate.

I have suggested that the common understanding concept embodies two epistemic requirements that must be added to standard Bayesian decision theory (Gintis 2009): mutual accessibility and symmetric reasoning. When common understanding obtains, players are able to develop a form of collective intentionality that shares many features with John Searle’s account. As a result, economists and in particular game theorists should not consider economic theory of institutions as competing with Searle’s theory of institutional facts. Indeed, the argument I have made in this paper suggests that game theory needs the Searlian perspective to provide a full explanation of institutions.

REFERENCES

- Aumann, Robert. 1976. Agreeing to disagree. *The Annals of Statistics*, 4 (6): 1236-1239.
 Aumann, Robert. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55 (1): 1-18.
 Aumann, Robert, and Adam Brandenburger. 1995. Epistemic conditions for Nash equilibrium. *Econometrica*, 63 (5): 1161-1180.

- Bacharach, Michael. 2006. *Beyond individual choice: teams and frames in game theory*. Princeton: Princeton University Press.
- Bardsley, Nicholas. 2006. On collective intentions: collective action in economics and philosophy. *Synthese*, 157 (2): 141-159.
- Binmore, Kenneth. 1998. *Just playing: game theory and the social contract II*. Cambridge (MA): MIT Press.
- Binmore, Kenneth. 2007. *Playing for real: a text on game theory*. Oxford: Oxford University Press.
- Binmore, Kenneth. 2009. *Rational decisions*. Princeton: Princeton University Press.
- Bowles, Samuel, and Herbert Gintis. 2011. *A cooperative species: human reciprocity and its evolution*. Princeton: Princeton University Press.
- Chwe, Michael S.-Y. 2003. *Rational ritual: culture, coordination, and common knowledge*. Princeton: Princeton University Press.
- Cubitt, Robin, and Robert Sugden. 2003. Common knowledge, salience and convention: a reconstruction of David Lewis' game theory. *Economics and Philosophy*, 19 (2): 175-210.
- Gintis, Herbert. 2009. *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton: Princeton University Press.
- Hakli, Raul, Kaarlo Miller, and Raimo Tuomela. 2010. Two kinds of we-reasoning. *Economics and Philosophy*, 26 (3): 291-320.
- Hédoin, Cyril. 2012. Linking institutions to economic performance: the necessity of macro-structures in micro-explanations. *Journal of Institutional Economics*, 8 (3): 327-349.
- Lewis, David. 2002 [1969]. *Convention: a philosophical study*. Oxford: Blackwell Publishing.
- Searle, John R. 1990. Collective intentions and actions. In *Intentions in communication*, eds. Philip R. Cohen, Jerry Morgan, and Martha E. Pollack. Cambridge (MA): MIT Press, 401-415.
- Searle, John R. 1995. *The construction of social reality*. London: Penguin Books.
- Searle, John R. 2010. *Making the social world: the structure of human civilization*. Oxford: Oxford University Press.
- Smit, J. P., Filip Buekens, and Stan du Plessis. 2011. What Is money? An alternative to Searle's institutional facts. *Economics and Philosophy*, 27 (1): 1-22.
- Sugden, Robert. 1991. Rational choice: a survey of contributions from economics and philosophy. *The Economic Journal*, 101 (407): 751-785.
- Sugden, Robert. 2000. Team preferences. *Economics and Philosophy*, 16 (2): 175-204.
- Sugden, Robert. 2003. The logic of team reasoning. *Philosophical Explorations*, 6 (3): 165-181.
- Tuomela, Raimo. 2007. *The philosophy of sociality*. Oxford: Oxford University Press.
- Tuomela, Raimo. 2011. Searle's new construction of social reality. *Analysis Reviews*, 71 (4): 707-719.
- Tuomela, Raimo, and Kaarlo Miller. 1988. We-intentions. *Philosophical Studies*, 53 (3): 367-389.
- Vanderschraaf, Peter. 1998. Knowledge, equilibrium, and convention. *Erkenntnis*, 49 (3): 337-369.

Cyril Hédoin is associate professor of economics at the University of Reims Champagne-Ardenne (France). His areas of specialization are in the philosophy of economics and institutional economics. He is currently working on the development of a non-individualistic understanding of game theory. He has published papers in the *Journal of Economic Issues*, the *Journal of Institutional Economics*, *Philosophy of the Social Sciences*, *Cahiers d'économie politique* (in French) and *OEconomia* (in French).

Contact e-mail: <cyril.hedoin@univ-reims.fr>

Early marginalist ideas on money: some neglected exceptions to the quantity theory

GERMÁN D. FELDMAN

National University of Quilmes

Abstract: The quantity theory of money (QTM) is an important building block of neoclassical economics. This has led scholars to believe that all monetary accounts proposed by marginalist economists are inherently based on the QTM. However, within the bimetallic controversy of the last quarter of the 19th century, there were some neoclassical proposals which departed from the framework of the QTM. In this article, I analyse three of these accounts: Alfred Marshall's symmetallism, Irving Fisher's compensated dollar plan, and Knut Wicksell's inconvertible paper standard. These monetary arrangements—especially the first two of them—have rarely been studied in the literature. Still, their relevance should not be neglected in current times in which the economics profession—both orthodox and heterodox approaches—has moved towards an endogenous money view. The proposals studied also show that the neutrality of money does not necessarily imply the QTM, as it is often suggested.

Keywords: bimetallic controversy, marginalism, monetary regimes, quantity theory, compensated dollar, symmetallism

JEL Classification: B13, E13, E31, E42

The establishment of a greater, and if possible absolute, stability in the value of money has thus become one of the most important practical objectives of political economy. But, unfortunately, little progress towards the solution of this problem has, so far, been made (Wicksell 1975 [1935], 7-8).

As can be appreciated in this quotation from a famous neoclassical economist, the first decades of marginalism in economics witnessed a

AUTHOR'S NOTE: I am grateful to Bertram Schefold, Luis Mireles-Flores, and two anonymous referees for helpful comments and suggestions. I also thank the financial support of the National Council of Scientific and Technical Research of Argentina (CONICET) and the German Academic Exchange Service (DAAD). Any remaining errors are my responsibility.

strong interest in discussions on monetary stability. Specifically, the objective to stabilise the average price level led scholars to postulate and debate alternative schemes to the prevailing gold standard system. Before the marginalist revolution, the notion of price stability in classical political economics consisted in keeping constant the purchasing power of money in terms of gold (or silver)—the standard of money under a metallic regime. Changes in the bullion price of commodities resulting from variations in mining costs were considered as “natural” and “inevitable” fluctuations (Ricardo 2004 [1810-1811], 65).

To deal with such short-run price fluctuations, the Peel’s Act (1844) was implemented in Britain. This Act divided the issuing and banking functions of the Bank of England into two independent bodies with the aim of bringing the gold standard to a sound basis according to the classical account. The Peel’s Act was meant as an application of the so-called ‘currency principle’ to rule out over-expansionary bias of monetary policy. According to the ‘currency principle’, in order to achieve a ‘sound’ currency, gold movements should be entirely reflected in domestic circulation, limiting thereby the capacity of the banking system to expand the money supply through credit operations.¹

After the marginalist revolution, monetary theorists—now equipped with a marginalist framework and holding a notion of price stability that implied the constancy of money prices—moved one step forward from the classical approach and tried to eliminate the remaining factor believed to induce permanent price fluctuations, the gold standard system.² A crucial difference between the classical and the marginalist views is that to the former, prices including those of precious metals are determined by the costs of production (leading to an *endogenous* view of money), while to the latter, income distribution and long-term prices are determined by supply and demand conditions (leading to an *exogenous* view of money).

The new marginalist proposals conferred a special role to the quantity theory of money (QTM). As is well known, in general equilibrium models relative prices, production levels, and ‘factor rentals’

¹ During the currency and banking schools debate that prevailed after the 1820s, the ‘currency principle’ was at the core of the dispute. The principle was strongly endorsed and defended by the school of Robert Torrens and Samuel Lloyd Jones, whereas it was opposed by the doctrine of Thomas Tooke and John Fullarton, who were ardent critics of the principle.

² In effect, the international gold standard proved unable to ensure price stability, as J. Laurence Laughlin (1909) documents for the last quarter of the 19th century, which registered periods of prolonged deflation and inflation.

(i.e., the real wage and the rate of interest) are firstly determined in the ‘real sector’ of the economy. This sector tends to full employment of resources, pushed by the substitution effect, and hence all nominal variables are established independently in the monetary sphere of the system once the quantity of money in circulation is specified. The QTM, though it originated much prior to the marginalist era, emerged as a suitable monetary-side counterpart to the neoclassical ‘barter-exchange’ description of market economies.

As a consequence of the use of the QTM as a key element in the neoclassical account, it is commonly believed that the founders of marginalism always argued on the basis of the QTM. For instance, Don Patinkin (1948; and 1965) states that the so-called ‘real-monetary dichotomy’ resided at the core of the contributions of Walras, Marshall, Wicksell, and basically most authors writing prior to the 1930s. Similarly, it is generally agreed that Marshall was a quantity theorist (see Humphrey 2004); and the prevailing view about Irving Fisher—primarily based on *The purchasing power of money* (1997 [1913])—is that he gave the QTM its modern shape (see Dimand 2000).

There are, however, three proposals endorsed by neoclassical authors that are exceptional in that they are founded on an endogenous view of money and are independent of the QTM, namely Alfred Marshall’s symmetallism, Irving Fisher’s compensated dollar, and Knut Wicksell’s inconvertible paper standard. These accounts have been seldom explored in the literature, with the possible exception of Wicksell’s contribution. Yet, a detailed analysis of these proposals is relevant for at least two reasons: First, the economics profession—both the orthodox and heterodox approaches—has moved towards an endogenous money perspective as a more realistic description of modern economies (which is commonly thought to be in itself a move away from neoclassical economics).³ Second, the study of these exceptional cases can establish new links between the marginalist view and the classical tradition, in which money was also conceived as an endogenous magnitude under a metallic or credit standard.

With the aim of exploring such monetary arrangements, this article is structured as follows. Section 1 discusses monetary regimes based on the validity of the quantity theory of money. The focus is placed on standard bimetallism, which represented the main rival to the gold

³ This trend can be appreciated, e.g., from the contributions to the new macroeconomic consensus or the post-Keynesian doctrine.

standard. Section 2 explores the three alternative schemes that emerged directly from the neoclassical school, but that are exceptional in that they departed from the QTM framework. The final section offers some concluding remarks.

MONETARY REGIMES ROOTED IN THE QUANTITY THEORY OF MONEY

The point of departure for the discussion of monetary regimes alternative to the gold standard can be traced to the ‘Bimetallic controversy’, which developed between the defenders of monometallic and those of bimetallic standards during the last quarter of the 19th century.⁴ At the empirical level, the experience of France, which at least *de jure* had successfully maintained a bimetallic system since the end of the Napoleonic wars to the 1870s,⁵ served as inspiration for pro-bimetallism. This view highlighted the stability of the relative price between gold and silver and of the exchange rate between countries with the gold standard and the silver standard during that period (especially between England and India).

In contrast to this steadiness observed in some instances, the last decades of the 19th century were characterised by a marked exchange rate instability and a generalised world deflation (the Great Depression).⁶ Its roots, in the opinion of bimetallicists, had to be found in the insufficient performance of the world gold supply when compared with a fast-growing bullion demand from economies experiencing a transition towards a gold standard and the rise of world aggregate output.⁷ In this section I present the classical qualms against

⁴ The Bimetallic controversy can be considered as the third big monetary controversy of the century, after the Bullionist controversy (i.e., the debates of the British Restriction Period (1797-1821) on the convenience of a commodity standard vis-à-vis inconvertible paper money), and the currency and banking schools’ controversy on the normal working of a convertibility system in the 1830s and 1840s.

⁵ The scope of bimetallicism was extended from France to its neighbours, when in 1865 the French government in company of Belgium, Switzerland, and Italy formed the Latin Monetary Union.

⁶ It should be noted that the term ‘Great Depression’ is not entirely adequate, since despite deflation, output and real wages continued to move upwards at “impressive rates” (Rostow 1947, 58).

⁷ In contrast, W. W. Rostow explains the declining trend of output prices by the rise of productivity in the British economy. The author also mentions, with a Kaleckian flavour, the possibility that lower profits and prices were explained by a decrease of the ‘degree of monopoly’ in the British economy after 1873 due to the rapid development of capital industries, among other factors, in the United States and Germany, and the existence of idle capacity when fixed investments matured, inducing domestic price-cutting competition (Rostow 1947, 233).

bimetallism in contrast to standard neoclassical proposals that were rooted in the QTM.

The classical view against bimetallism

Two opposite views on the long-period viability of *standard* bimetallism can be found in the literature. On the one hand, the classical school,⁸ endorsing a cost-of-production determination of the relative price of gold in terms of silver, considered that bimetallism was intrinsically unstable. Specifically, it faced a ‘knife-edge’ problem: if the official parity did not match the market ratio, the metal over-valued in the market would disappear from circulation, leading to a *de facto* mono-metallic standard.⁹ In effect, as David Ricardo expresses:

No permanent measure of value can be said to exist in any nation while the circulating medium consists of two metals, because they are constantly subject to vary in value with respect to each other. However exact the conductors of the mint may be, in proportioning the relative value of gold to silver in the coins, at the time when they fix the ratio, they cannot prevent one of these metals from rising, while the other remains stationary, or falls in value. Whenever this happens, one of the coins will be melted to be sold for the other. Mr. Locke, Lord Liverpool, and many other writers, have ably considered this subject and have all agreed, that the only remedy for the evils in the currency proceeding from this source, is in making one of the metals only the standard of value (Ricardo 2004 [1810-1811], 19-20; italics added)

Therefore, to classical authors, a bimetallic regime *at a fixed parity*¹⁰ would follow an unstable path affected by the operation of Gresham’s law, once market conditions have changed from the initial levels used as a reference for setting the official parity.

Let us use a simple example presented by Laughlin (1896, 26) to understand its functioning. Initially, the mint and market ratio of gold

⁸ For a description of the classical approach articulated along Sraffian lines, see Garegnani 1984; 1990; and Eatwell 1977.

⁹ Of course, the market ratio may permanently depart from the mint ratio between the boundaries imposed by the ‘gold-silver’ points in a similar way to the operation of a gold standard.

¹⁰ Once symmetallism is introduced into the picture, the references to ‘standard’ bimetallism or bimetalism ‘at a fixed parity’ will become much clearer. Put it shortly, the conventional approach to a double standard implies arbitrary fixing the parity between gold and silver, while an alternative option could be to define the quantity ratio or the composition of coins, while leaving the value ratio to be set by market conditions.

relatively to silver are both 1:15, but market conditions change and the new relative price becomes 1:16. That is, in the market it requires sixteen ounces of silver to buy one ounce of gold bullion; but at the Mint the Government receives fifteen ounces of silver, and coins it into silver coins which are legally equivalent to one ounce of gold. The holder of silver will be tempted to bring his silver to the Mint and exchange it for silver coins. The money brokers, who are better informed than the general public about the value of metals, lead the arbitrage process. They exchange the silver coins obtained from the Mint for gold coins as long as gold coins remain in circulation. Having received an ounce of gold in coin for their fifteen ounces of silver coin, they can at once sell the gold as bullion (after melting it, or selling it to exporters) for sixteen ounces of silver bullion. They retain one ounce of silver as profit, and with the remaining fifteen ounces of silver go to the Mint for more silver coins, repeating the whole process. And thus, as Laughlin puts it:

The existence of a profit in selling gold coins as bullion, and presenting silver to be coined at the Mint, is due to the divergence of the market from the legal ratio, and no power of the Government can prevent one metal from going out of circulation (1896, 27).

In a similar vein, T. Lloyd (1894, 35) argues that during its seventy years of formal existence, French bimetallism was never operative. Both gold and silver never circulated at the same time. Up to 1848 France was a silver-using country; afterwards it became gold-using.¹¹ And Willis (1895) provides evidence of Gresham's law indeed operating in France during the bimetallic period.

The marginalist view in favour of bimetallism

Early marginalist authors such as Walras (1977 [1874]), Marshall (1923 [1887]) and Fisher (1894),¹² defended bimetallism based on the validity

¹¹ However, Lloyd remarks the positive effects of bimetallism, which, though not sustainable, was a natural transition from a silver standard, which characterised European countries before the 19th century, to the gold standard, which dominated by the end of the 19th century (Lloyd 1894, 31-32).

¹² As an exception, Jevons was a declared gold monometallist who held an optimistic view of the responsiveness of world gold supply and stressed the lower bulk of gold as money vis-à-vis silver, especially for countries experiencing a growing wealth. It is quite remarkable that the author—though sharing with Walras and Marshall a theory of value and distribution based on marginal utility—remained much closer to the classical tradition with regard to the treatment of the value of money under a commodity standard: “the value of gold and silver are ultimately governed, like those of all commodities, by the cost of production” (Jevons 1886 [1881], 100). Jevons, like

of the quantity theory of money (see Laidler 2004). In contrast to the use given to it by classical authors, the QTM was then not only applied to the case of inconvertible paper money, but also to a metallic circulation. Neoclassical economists believed that the instauration of an official parity between gold and silver did not imply replacing a ‘natural’ value with an ‘artificial’ discretionary ratio, but filling in a gap left by the market due to the specificities of precious metals:

[...] gold and silver have no natural value. They are so durable that the year’s supply is never more than a small part of the total stock, and therefore their values do not conform closely to their costs of production. And, in so far as their values are regulated by the relations between the demands for them and the existing stocks of them, their value is artificial, because the demand for them as currency is itself artificial (Marshall 1923 [1887], 200-201).

Fisher’s exposition of a bimetallic standard allows comparing the classical and neoclassical views on the subject in order to establish theoretical and methodological differences. Fisher stresses that market prices are slow in their gravitation towards normal prices in the case of precious metals due to the slow adjustment of supply. As a result, two different equilibria arise: a ‘temporary equilibrium’, with a given stock of precious metals, and a ‘normal or permanent equilibrium’, with an endogenous stock of gold and silver (see Fisher 1997 [1913], 124).

To see this more clearly, consider the following general equilibrium model. Let us assume an economy with $n > 2$ commodities, including gold (g) and silver (s). Land is free and joint production is excluded. Commodity j (with $j \neq g, s$) is employed as the *numéraire* of the price system ($p_j = 1$). Long-period or normal prices for consumption and

Ricardo and his followers, employed the quantity theory only as a short-run adjustment mechanism. In the long period, “the common argument that there will not be enough gold to carry on the trade of the world with, does not stand a moments’ examination in this aspect. In the first place, if the value of gold rises, more gold will be produced, and the great number of gold-mining enterprises now being put forth may have some connection with this principle” (Jevons 1886 [1881], 100). To which he adds: “When we turn to the *temporary* view of the subject, by which I mean the circumstances and interests of the next ten or fifteen years, the difficulties increase, chiefly because the data become wholly uncertain and contingent. The great principle of the cost of production fails us, because in the case of such durable commodities as gold and silver, the accumulated stock in hand is immensely greater than the annual production or consumption. It stands to reason, of course, that if several great nations suddenly decide that they will at all cost have gold currencies to be coined in the next few years, the annual production cannot meet the demand, which must be mainly supplied, if at all, out of stock” (Jevons 1886 [1881], 102).

capital goods (i.e., excluding precious metals) are defined by the following ($n-2$) conditions:

$$p_{i/j} = c_i^{13} \quad (1)$$

The quantity system includes n market-clearing conditions:¹⁴

$$Q_i^d(p_{i/j}, \dots, p_{n/j}) = Q_i^s(p_{i/j}, \dots, p_{n/j}) \quad (2)$$

Let us note that the demands for gold and silver emerging from the general equilibrium system only reflect their industrial or non-monetary uses (Q_{gb}^d and Q_{sb}^d). The quantity of gold and silver circulating as money will therefore result from deducing the industrial demand from the *given* stock of precious metals (\bar{Q}_g^s , \bar{Q}_s^s):

$$\begin{aligned} Q_g^d &\equiv Q_{gb}^d + Q_{gM} = \bar{Q}_g^s \\ Q_s^d &\equiv Q_{sb}^d + Q_{sM} = \bar{Q}_s^s \end{aligned} \quad (3)$$

As has been previously mentioned, the implementation of a bimetallic system implies that some proportion of both metals must circulate simultaneously as money. Therefore, the following two restrictions must be added:

$$\begin{aligned} Q_{gM} &= Q_g - Q_{gb} > 0 \\ Q_{sM} &= Q_s - Q_{sb} > 0 \end{aligned} \quad (4)$$

Finally, monetary prices will result from the ‘exchange equation’, which reflects that the value of transactions performed in a certain period must necessarily be sustained by an equivalent amount of circulating medium:

¹³ In this expression, c_i is defined by the normal cost of production of commodity i , which reflects the dominant technique and the state of income distribution. Assuming constant returns to scale demand conditions will influence natural prices only through their effect on normal distribution.

¹⁴ For simplicity, the market clearing conditions for the factors of production (labour and capital) are omitted from the analysis, but a complete version of the general equilibrium system should include them in order to find the equilibrium values of the wage rate and the rate of interest (profits).

$$\frac{1}{v} \left(\sum_{i=1}^{n-2} p_{i/j} Q_i \right) p_{j/g} = Q_{gM} + p_{s/g} Q_{sM} \quad (5)$$

Where:

v = ‘normal’¹⁵ velocity of circulation

$p_{s/g}$ = price of silver in terms of gold

As can be seen from (5), the ‘exchange equation’ determines the gold price of commodities ($p_{j/g}$), once the relative price between gold and silver is set ‘from outside’,¹⁶ in line with the QTM, whose central elements are an *exogenous* money supply and an *endogenous* price level.

There was a general consensus among the advocates of bimetallism on the fact that for a small open economy, viz. one that faced a given gold-silver ratio in the rest of the world, bimetallism at an official parity that is different from the international parity could not be permanently maintained. This is so because of the operation of arbitrageurs exploiting the profits of bullion trade. The emphasis of pro-bimetallists on international cooperation in monetary matters can therefore be easily grasped:¹⁷

But let us suppose for the sake of argument ‘one metal or the other’ to be the dearer of the two, and that therefore it ‘will leave the country’, and the depreciated metal will remain. It is admitted on all

¹⁵ As Mauro Caminati (1981) suggests, important changes in the composition of demand for monetary assets tended to be confined to periods of alarm when the state of confidence collapsed. In normal times, the ratio of the demand for gold and bank-notes to monetary income tended to be relatively stable.

¹⁶ Neither Walras nor Fisher believed that the gold-silver ratio could be freely set disregarding market conditions derived from their uses in art and industry. Indeed, there was a range of values in which bimetallism was possible, defined by the relative prices that would be effective under monometallic gold and silver standards. “The legislator may undoubtedly announce an arbitrary ratio of value between gold and silver; but what he cannot do is ensure that this ratio is kept up, or even that it will become established if it deviates too far from a certain value in line with the actual circumstances. If this legal value were too high, i.e., too much to the advantage of gold, then all silver would remain in the form of merchandise and, in fact, the legislator would decree gold monometallism; if it were too low, i.e., too much to the advantage of silver, then gold would remain in the form of merchandise and, in fact, the legislator would have decreed silver monometallism” (Walras 2005 [1884], 117).

¹⁷ The defenders of French bimetallism thus considered that France, though surrounded by economies with other gold-silver ratios, concentrated a high proportion of the world stock of precious metals, so that its domestic conditions could influence their international values.

hands that such was the case with France when she stood alone a bimetallic nation surrounded by monometallic neighbours; and that such must be the case with any single country so standing. It must with equal unanimity be admitted that if *all* commercial nations without exception had but one mind in the matter and received Gold and Silver alike (in a fixed proportion) in payment of debt, no such exodus of one or other could take place. Whither would the dearer metal go? Surely the objectors will not say that it will leave all countries simultaneously! (Gibbs 1886 [1881], 40).

Nevertheless, it is clear—from a marginalist perspective—that a closed economy with gold and silver mining, as well as the world economy as a whole, could implement bimetallism without major inconveniences. To Gibbs, gold and silver in these cases are in fact two complementary parts of the same whole:

Under such a compact, Silver and Gold are as one metal-limbs of the same body, parts of the same whole-fused like an electrum into one mass; and when both are recognized as the measure of other commodities, whose value as a mass varies with the total quantity of that measure, the cost of production does not practically determine the mutual value of the two parts of it (Gibbs 1886 [1881], 39).

To sum up, a critique of bimetallism as a permanent monetary regime should call into question the validity of a double standard within the boundaries of a closed economy. The ‘market ratio’ should have a meaning beyond the official parity in different countries. The key issue in relation to the feasibility of bimetallism is the role of the exchange ratio between gold and silver defined by costs equations. Only if the flow of annual production and thus the costs of production are placed at the core of the supply side of gold and silver, can bimetallism be fundamentally criticized. Only if bullion mining yields the average rate of profits, can productive capitals be directed to it and a regular flow of production necessary for reproducing the system be granted. If the mint ratio, for instance, undervalues silver, then capital will leave silver mining and flow to other branches of production where a higher rate of profits can be obtained. The fundamental flaw of standard bimetallism would be thus the attempt to fix two *numéraires*, hence over-determining the normal price system.¹⁸

¹⁸ As Piero Sraffa (1960) masterfully shows, once one distributive variable is specified, there is only room to set one *numéraire* (this fact does not rule out the possibility of choosing the price of a bundle of commodities).

Fisher's solution and Walras's alternative

Following Pierangelo Garegnani (1976), one could assimilate the normal equilibrium of Fisher to the long-period equilibrium of classical economists.¹⁹ Under a permanent equilibrium, the cheap metal pushes the dear one out of circulation, destroying bimetallism as Ricardo formerly suggested. However, according to Fisher, the physical properties of precious metals and the specificities of their supply make temporary and normal equilibriums quite distinctly separated in time:

The time of redistributing existing stocks of metal, according to a newly enacted law, depends on the rapidity of transportation, melting, and minting, and would be measured in months or weeks. Normal equilibrium, however, depends on the slow working of changes in the rates of production and consumption, and would be measured in years (Fisher 1997 [1913], 124)

In the meantime, counterbalancing forces may arise, such as changes in the conditions of production of the dear metal, so as to revert the natural price towards the official parity. In other words, while the normal equilibrium could be relevant in theoretical terms, its existence being enough to challenge the viability of bimetallism, it could be disregarded in practical terms, since before converging to such a position the economy may adopt a different path.

A further illustration of a monetary regime relying on the QTM is Walras's *gold standard with restrictive minting of silver* (2005 [1884]). Such a scheme represented an intermediate position between a double standard and gold monometallism. On the one hand, this system held with bimetallism the coexistence of gold and silver as circulating media. On the other, it shared with the gold standard the fact that gold alone had a fixed price, while silver was a mere token, with its purchasing power as a coin being higher than its value as a commodity. Under this arrangement, the government would be able to expand or contract the money supply in order to stabilize the price level. In Walras's terms:

The State will increase or decrease its quantity according to the circumstances. If the *rareté* and, consequently, the value of money

¹⁹ Garegnani argues that neoclassical theory until the mid-1930s shared with classical economics the implied notion of equilibrium, which was understood as a long period position in which prices are such so as to allow a uniform rate of profit over supply prices.

tend to climb over the limit assigned to it, the quantity of the special token should be increased. This would allow for the demonetization of a certain quantity of gold, reducing the *rareté* and the value of the money commodity. If the *rareté* and, consequently, the value of money tended to remain below the limit, the quantity of special token should be lowered. This would entail the monetization of a certain quantity of gold and increase the *rareté* and value of the money commodity (Walras 2005 [1884], 8).

In this respect, over-valued silver would not displace gold from circulation, because the mint would not be open to the free coinage of silver. Thus the government would be the agent in charge of buying silver bullion in the market at the prevailing price and subsequently coining the metal, rather than arbitrage decisions of the private sector.

NEOCLASSICAL REGIMES NOT BASED ON THE QUANTITY THEORY OF MONEY

There exist some neglected exceptions among neoclassical scholars to the usual recourse to the QTM when exploring commodity standards. Although it may seem surprising to some at first glance, monetary regimes that are independent of the QTM are a possibility within the marginalist framework. This is so because a general equilibrium system is compatible with an endogenous stock of precious metals and a natural value of bullion defined by the costs of production, rather than relative scarcity.²⁰

In terms of the general equilibrium model described in the previous section, two relevant changes have to be made. First, the system of normal prices should include the cost-of-production equations for gold and silver, resembling those of other reproducible commodities. In this case gold is taken as the *numéraire*:

$$\begin{aligned} p_{i/g} &= c_i \quad (i = 1 \dots n-1) \\ 1 &= c_g \end{aligned} \quad (1')$$

Secondly, since the relative price between gold and silver ($p_{s/g}$) comes to be defined by structural conditions in (1'),²¹ the 'exchange equation' will be interpreted now as determining the equilibrium quantity of money. As the new interpretation of equation (5) reveals, an additional condition must be introduced in order to define the monetary

²⁰ See Jürgen Niehans 1978, chapter 8.

²¹ See footnote 13.

demand for each precious metal under a bimetallic standard. Specifically:

$$Q_{gM} = \bar{k}Q_{sM} \quad (6)$$

With k an arbitrary constant setting the composition of the money stock. As will be discussed in the following subsection, there is an alternative to bimetallism proposed by Alfred Marshall, called symmetallism, which precisely provides the required condition.

Marshall's symmetallism

The symmetallism system exposed by Alfred Marshall in his essay “Remedies for fluctuations in general prices” (1923 [1887])—*which according to the* British scholar found inspiration in Ricardo’s Ingot plan—represents a variant of a bimetallic standard that fixes the quantity ratio, instead of the value ratio. This is how he describes it:

I propose that currency should be exchangeable at the Mint or Issue Department not for gold, but for gold and silver, at the rate of not £1 for 113 grains of gold, but £1 for 56 ½ grains of gold, together with, say, twenty times as many grains of silver. I would make up the gold and silver bars in gramme weights, so as to be useful for international trade. A gold bar of 100 grammes, together with a silver bar, say, twenty times as heavy, would be exchangeable at the Issue Department for an amount of the currency which would be calculated and fixed once for all when the scheme was introduced. (It would be about £28 or £30 according to the basis of calculation.) Any one who wanted to buy or sell gold or silver alone in exchange for currency could get what he wanted by exchanging gold for silver, or silver for gold, at the market rate. Government fixing its own rates from day to day, so as to keep its reserves of the two metals in about the right proportion, might safely undertake this exchange itself; and then any one could buy or sell either gold or silver for currency in one operation (Marshall 1923 [1887], 204-205).

At the time of the introduction of this proposal, it was presented as a variant *but* equivalent alternative to the typical bimetallism with a fixed value ratio.²² Nevertheless, the fact that symmetallism does not aim at imposing an arbitrary relative price makes this arrangement

²² Indeed, as Walras (1977 [1874], 339) shows, in a general equilibrium system with a bimetallic monetary standard, there are n independent equations to solve for $n+1$ variables, so that there exists one degree of freedom.

potentially compatible with the classical approach. In effect, a ‘classical’ bimetallic standard would imply a metallic circulation of coins whose technique of production employed fixed quantities of gold and silver as inputs, or a circulation of paper money convertible on demand into certain quantities of gold *and* silver bullion:

$$p_{m/g} = (b_m^g + b_m^s p_{s/g}) = 1 \quad (7)$$

Where:

$p_{m/g}$ = value of a coin (m) in terms of gold bullion

b_m^g = quantity of gold bullion necessary to produce one coin

b_m^s = quantity of silver bullion necessary to produce one coin

The purchasing power of such money would be defined by its respective cost of production. In addition, the symmetallic standard would be more stable than a monometallic standard, because the depreciated metal would erode the value of money only by its weight on the composite commodity.²³ At the same time, the value of the coin would be subject to more regular fluctuations for it would change with every change in value of either of the two metals, instead of only one (see Miller 1898, 276).

Indeed if the bimetallists had changed two single words in their formulation they would have been proposing something better than the gold standard instead of something worse. If instead of saying, “Let a dollar stand for 25.8 grains of gold *or* 412.5 grains of silver” they had said, “Let two dollars stand for 25.8 grains of gold *and* 412.5 grains of silver”, and if this proposal had been adopted, then some of the more serious fluctuations in prices could have been avoided (see Lewis 1925, 42).

In other words, Marshall’s proposal aims to ensure a double standard, but avoid the intrinsic instabilities connected with the attempt to arbitrarily fix a relative price between two commodities.

²³ Marshall’s symmetallicism was later extrapolated into schemes that included a larger number of commodities. For instance, Benjamin Graham (1937) proposed that the dollar should be defined in terms of a fixed-weight basket of 23 commodities, and that the Federal Reserve issue notes against warehouse receipts for the basket thus established. He selected his commodities on the strength both of their economic importance and their storability. Simultaneously but independently, Frank Graham (1949) advocated a similar plan.

Fisher's compensated dollar

Irving Fisher's plan for a 'compensated dollar' was confined to a gold standard system and therefore, was less radical than Marshall's proposal, in that it implied continuity with the most common monetary regime of the time. It intended to neutralise changes in the gold price of commodities by counterbalancing variations in the official price of gold, thus leaving the general price level as constant. This is Fisher's presentation of his proposal:

The plan aims to make the purchasing power of the dollar constant. It would compensate for any loss of purchasing power of each grain of gold by increasing the number of grains which go to make a dollar. In other words it aims to standardize the dollar as a unit of purchasing power. We have standardized the yard, the pound, the kilowatt, and every other important commercial unit except the most important of all, the dollar, the unit of purchasing power. We have now a gold dollar of constant weight, but of varying purchasing power. We need a dollar of constant purchasing power and varying weight (Fisher 1997 [1913], 214).

In terms of implementation, Fisher planned to manipulate the seigniorage rate on gold coinage ('*seg*' hereafter), readjusting it according to changes in the general price level. As commodity prices rose, the weight of the 'bullion dollar' or the normal value of money would be sustained by a rise of *seg*. "The increasing number of grains of bullion going to make a dollar would then compensate for the lessening purchasing power of each grain" (Fisher 1997 [1913], 218). Analytically:

$$\begin{aligned} \bar{p}_{i/m} &= \uparrow p_{i/g} \cdot \downarrow p_{g/m} \\ \uparrow p_{m/g} &= b_m (1 + \uparrow seg) \end{aligned} \quad (8)$$

One possible objection to this arrangement is how to deal with a situation of falling prices instead of rising ones. Though a detailed analysis of the working of coinage is beyond the scope of the present inquiry, we know that the gold price of coins ($p_{m/g}$) cannot fall below its natural level for an amount higher than the melting cost, for then all the gold coin will at once be melted into bullion, in which form it will be worth more than as coin. This fact is recognised by Fisher himself, who expresses that: "In a period of rising prices, regulation would be easy; in a period of falling prices, regulation might be quite impossible"

(Fisher 1997 [1913], 331). In addition, John Maynard Keynes (1971 [1926], 126) considers that Fisher's plan implied "a preference for stability of internal price level over stability of external exchange". Furthermore, to periodically change the mint price of gold in order to offset fluctuations in the value of bullion would imply "the ultimate abandonment of gold as a monetary standard" (Cassel 1920, 43).²⁴

The lack of any role for the QTM in Fisher's programme is explicitly recognized by the American scholar in a newspaper article: "there is nothing whatever in the plan to standardize the dollar which could not be accepted by those who reject the quantity theory altogether" (Fisher, *New York Times*, December 22, 1912). This fact astonished modern neoclassical authors such as Don Patinkin, who, far from attempting a logical critique of the proposal, provided instead an ad-hoc justification:

[...] the person who is our present concern is not Irving Fisher the author of the scientific work on *The Purchasing Power of Money*, but Irving Fisher the deviser of a plan to be 'sold' to the economics profession as well as to the business community and government—and to be 'packaged' accordingly. The quantity theory of money was out of favour in some circles, so the plan should not be explicitly associated with it. The commodity theory of money had influential supporters, so the plan should be presented in language that had the sounds of that theory. The gold standard was sacred, so it should be emphasized that the plan did not involve its abandonment (Patinkin 1993, 9).²⁵

To sum up, Fisher's plan for a monetary regime (which was not rooted in the QTM) was essentially intended to offset the fluctuations in the value of gold with counterbalancing movements in the purchasing power of the unit of account.

Wicksell's inconvertible paper standard

As a step towards eliminating the long-period fluctuations of money prices induced by changes in the productivity of gold mines, Knut Wicksell advocated divorcing the monetary standard from precious

²⁴ The immediate consequence of achieving price stability would be the variability of the money price of gold. Yet, it should be noticed that the periodic redefinition of a fixed standard cannot be put on the same grounds of a fully flexible monetary regime.

²⁵ Interestingly, contemporary critics such as Frank Taussig (1913, 402) argue precisely the opposite, namely, that Fisher's proposal was designed to restrain the money supply and thus, to stabilize the general price level in a standard quantity-theoretic basis. In order to achieve such result, Taussig interprets the quantity of gold retained by the State as seigniorage as a reduction in the quantity of money in circulation.

metals and from any other commodity, by introducing instead a 'managed' inconvertible paper currency. This is how Wicksell presents the significance of his proposal:

Only by completely divorcing the value of money from metal, or at any rate from its commodity function, by abolishing all free minting, and by making the minted coin or banknotes proper, or more generally the unit employed in the accounts of the credit institutions, both the medium of exchange and the measure of value—only in this way can the contradiction be overcome and the imperfection be remedied. It is only in this way that a logically coherent credit system, combining both economy of monetary media and stability in the standard of value, becomes in any way conceivable (Wicksell 1975 [1935], 126).

Compared with the previous marginalist scholars, Wicksell had a far better understanding of the endogenous nature of the money supply under a credit-based monetary regime. As a result, his position regarding the validity of the quantity-theoretic use of the 'exchange equation' was closer to the classical banking school, which stated that:

Bank-notes [...] are never issued but on the demand of the recipient parties. New gold coin and new conventional notes are introduced into the market by being made the medium of *payments*. Bank notes, on the contrary, are never issued but on *loan*, and an equal amount of notes must be returned into the bank whenever the loan becomes due. Bank-notes never, therefore, can clog the market by their redundancy, nor afford a motive to any one to pay them away at a reduced value in order to get rid of them. The banker has only to take care that they are lent on sufficient security, and the reflux and the issue will, in the long run, *always* balance each other (Fullarton 1845, 64; italics added).

The main aim of Wicksell was to find a way to stabilize the price level without making recourse to an exogenous money supply. In this regard, he focused directly on the effective demand, identifying the money rate of interest as the instrument through which the banking system influenced the general price level. This influence is triggered by creating a gap between aggregate supply and demand in the market for commodities, which corresponds to a divergence between the money rate of interest and the 'natural' rate of return on capital. The focus on the interest rate was stressed by Wicksell as follows:

There is a certain rate of interest on loans which is neutral in respect to commodity prices, and tends neither to raise nor to lower them. This is necessarily the same as the rate of interest which would be determined by supply and demand if no use were made of money and all lending were effected in the form of real capital goods. It comes to much the same thing to describe it as the current value of the *natural rate of interest on capital* (Wicksell 1936 [1898], 102; italics in the original).

Wicksell's method of stabilizing the purchasing power of money was straightforward: the banking system, which possessed the ability to set the money rate of interest, had to adjust the rate to the natural level defined by productivity and thrift. Let us suppose that banks lend money at a rate of interest different from that which corresponds to the current value of the natural rate on capital. In this case,

The economic equilibrium of the system is *ipso facto* disturbed. If prices remain unchanged, entrepreneurs will in the first instance obtain a surplus profit (at the cost of the capitalists) over and above their real entrepreneur profit or wage. This will continue to accrue so long as the rate of interest remains in the same relative position [...] And the number of people becoming entrepreneurs will be abnormally increased. As a consequence, the demand for services, raw materials, and goods in general will be increased, and the prices of commodities must rise" (Wicksell 1936 [1898], 105-106; italics in the original).

Likewise, if there is an increase in the rate of interest, the opposite situation would occur. Entrepreneurs would suffer losses below their normal incomes as long as prices are stable, hence there would be an incentive to move investment towards more profitable markets. As a consequence there would be an excess supply for goods and services, and the price level would fall (see Wicksell 1936 [1898], 106).

Wicksell's conclusion then was that a paper standard, when administered efficiently, would be superior to any commodity standard in terms of price stability. And to achieve an efficient monetary administration under this regime, the main challenge of monetary policy should be to find the natural level of the interest rate. Wicksell's proposal is clearly an alternative to proposals rooted in the QTM, since it is intended to be independent of any attempt to affect the level of prices through exogenous changes in the quantity of money in the economy.

CONCLUDING REMARKS

During the last decades of the 19th century and the beginning of the 20th century, neoclassical economics centred on the question of how to stabilize the price level. Most authors advocated bimetallism, whose theoretical support resided in the validity of the QTM, an account that closed the nominal system with complete independence from general equilibrium barter-exchange models describing the real economy. Yet, contrary to the common belief found in the literature, there are neoclassical authors who proposed alternatives to bimetallism that are independent of the QTM, and which are based on an endogenous view of money. The endogenous dynamics makes these alternative approaches closer to classical views, rather than to the standard neoclassical exogenous view.

In particular, the scope of Marshall's symmetallism is much broader than just a plausible alternative for standard bimetallism. It is the only version of a double standard that could be consistently regarded as stable from a long-period perspective. Alternatively, Fisher's proposal of a compensated dollar is fully compatible with the classical view on commodity standards with seigniorage, and it is fairly independent of the QTM. Finally, Wicksell's approach presents several coincidences with the endogenous money view of the classical banking school. One merit of Wicksell is that he changed the focus from the quantity of money to the rate of interest as the main instrument of monetary policy. A corollary of the analysis presented in this article is that the neutrality of money does not necessarily imply the QTM, as frequently suggested.

REFERENCES

- Caminati, Mauro. 1981. The theory of interest in the classical economists. *Metroeconomica*, 33 (1-2-3): 79-104.
- Cassel, Gustav. 1920. Further observations on the world's monetary problem. *The Economic Journal*, 30 (117): 39-45.
- Dimand, Robert. 2000. Irving Fisher and the quantity theory of money: the last phase. *Journal of the History of Economic Thought*, 22 (3): 329-348.
- Eatwell, John. 1977. The irrelevance of returns to scale in Sraffa's analysis. *Journal of Economic Literature*, 15 (1): 61-68.
- Fisher, Irving. 1894. The mechanics of bimetallism. *The Economic Journal*, 4 (15): 527-537.
- Fisher, Irving. 1997 [1913]. The purchasing power of money. In *The works of Irving Fisher*, ed. William J. Barber. London: Pickering and Chatto.
- Fullarton, John. 1845. *On the regulation of currencies*. London: John Murray.

- Jevons, William Stanley. 1886 [1881]. Bimetallism. *Contemporary Review*, May. Reprinted in *The bimetallic controversy*, eds. H. Gibbs, and R. Grenfell. London: Effingham Wilson, Royal Exchange.
- Garegnani, Pierangelo. 1976. On a change in the notion of equilibrium in recent work on value and distribution. In *Essays in modern capital theory*, eds. M. Brown, K. Sato, and P. Zarembka. Amsterdam: North Holland Publishing Company.
- Garegnani, Pierangelo. 1984. Value and distribution in the classical economists and Marx. *Oxford Economic Papers*, 36 (2): 291-325.
- Garegnani, Pierangelo. 1990. Quantity of capital. In *The New Palgrave: Capital Theory*, eds. J. Eatwell, M. Milgate, and P. Newman. London: W. W. Norton & Co.
- Graham, Benjamin. 1937. *Storage and stability*. New York: McGraw-Hill.
- Graham, Frank. 1949. *Social goals and economic institutions*. New Jersey: Princeton University Press.
- Gibbs, Henry H. 1886 [1881]. The double standard. In *The bimetallic controversy*, eds. H. Gibbs, and R. Grenfell. London: Effingham Wilson, Royal Exchange.
- Humphrey, Thomas. 2004. Alfred Marshall and the quantity theory of money. *Federal Reserve Bank of Richmond, Working Paper No. 04-10*. Richmond, VA.
- Keynes, John Maynard. 1971 [1926]. A tract on monetary reform. In *The collected writings of John Maynard Keynes, vol. IV*. London: Macmillan.
- Laidler, David. 2004. From bimetallism to monetarism: the shifting political affiliation of the quantity theory. In *Political events and economic ideas*, eds. I. Barends, V. Caspari, and B. Schefold. Cheltenham: Edward Elgar.
- Laughlin, J. Laurence. 1896. *The history of bimetallism in the United States*. New York: D. Appleton & Co.
- Laughlin, J. Laurence. 1909. Gold and prices, 1890-1907. *The Journal of Political Economy*, 17 (5): 257-271.
- Lewis, Gilbert. 1925. A plan for stabilising prices. *The Economic Journal*, 35 (137): 40-46.
- Lloyd, T. 1894. *Bimetallism examined*. London: Statist Press.
- Marshall Alfred. 1923 [1887]. Remedies for fluctuations in general prices. In *Memorials of Alfred Marshall*, ed. Arthur C. Pigou. London: Macmillan.
- Niehans, Jürgen. 1978. *The theory of money*. Baltimore and London: The Johns Hopkins University Press.
- Patinkin, Don. 1948. Relative prices, Say's law, and the demand for money. *Econometrica*, 16 (2): 135-154.
- Patinkin, Don. 1965. *Money, interest, and prices: an integration of monetary and value theory*, 2nd ed. [abridged]. Cambridge (MA): The MIT Press.
- Patinkin, Don. 1993. Irving Fisher and his compensated dollar plan. *Federal Reserve Bank of Richmond Economic Quarterly*, 79 (3): 1-34.
- Ricardo, David. 2004 [1810-1811]. High price of bullion: a proof of the depreciation of bank notes. In *The works and correspondence of David Ricardo, vol. IV*, ed. Piero Sraffa. Indianapolis: Liberty Fund.
- Rostow, W. W. 1947. *British economy of the nineteenth century*. Oxford: Clarendon Press.
- Sraffa, Piero. 1960. *Production of commodities by means of commodities*. Cambridge (UK): Cambridge University Press.

- Taussig, Frank. W. 1913. The plan for a compensated dollar. *The Quarterly Journal of Economics*, 27 (3): 401-416.
- Walras, León. 1977 [1874]. *Elements of pure economics*, translated by William Jaffe. New Jersey: Augustus M. Kelly.
- Walras, León. 2005 [1884]. *Studies in applied economics: theory of the production of social wealth*, vol. 1. London: Routledge.
- Wicksell, Knut. 1936 [1898]. *Interest and prices*. New York: Sentry Press.
- Wicksell, Knut. 1975 [1935]. *Lectures on political economy*, vol. II. New Jersey: Augustus M. Kelley.

Germán D. Feldman received his doctoral degree in economics from the Goethe University Frankfurt. He is professor of macroeconomics at the National University of Quilmes and works as advisor to the Under-Secretary of Economic Policy at the Ministry of Economy and Finance of Argentina. His research focuses on monetary theory and policy, classical economics, and history of economic thought.
Contact e-mail: <feldmangerman@gmail.com>

By “fancy or agreement”: Locke’s theory of money and the justice of the global monetary system

LUCA J. UBERTI

American University in Kosovo
Rochester Institute of Technology

Abstract: Locke argues that the consent of market participants to the introduction of money justifies the economic inequalities resulting from monetarization. This paper shows that Locke’s argument fails to justify such inequalities. My critique proceeds in two parts. Regarding the consequences of the consent to money, neo-Lockeans wrongly take consent to justify inequalities in the original appropriation of land. In contrast, I defend the view that consent can only justify inequalities resulting directly from monetized commercial exchange. Secondly, regarding the nature of consent, neo-Lockeans uncritically accept Locke’s account of money as a natural institution. In contrast, I argue that money is an irreducibly political institution and that monetary economies cannot develop in the state of nature. My political account of money has far-reaching implications for the normative analysis of the global monetary system and the justification of the economic inequalities consequent upon it.

Keywords: Locke, metallism, chartalism, consent, equality, global monetary system

JEL Classification: B11, E42, Z10

The extent to which Locke’s principles of justice (or ‘Law of Nature’) justifies permissible material inequalities is a long-standing terrain of contention in the neo-Lockean tradition. While some critics (e.g., left-libertarians) have argued that the way Locke’s law regulates original resource appropriation contains extensive egalitarian provisions (Vallentyne, et al. 2000), little thinking has been devoted to the problem of theorizing the inequalities resulting from resource transfer and commercial exchange. This is unfortunate, for there is reason to think

that the *problématique* that Robert Nozick (1974, 150) calls "justice in transfer" was high up on Locke's politico-philosophical agenda. In particular, a significant portion of chapter 5 of the *Second treatise of government* is devoted to justifying the inequalities that emerge upon the introduction of money. Here, Locke makes what in contemporary political theory would qualify as a hard-headed libertarian argument:

[s]ince gold [...] has its value only from the *consent* of Men [...], it is plain, that Men have agreed to a disproportionate and *unequal* Possessions of the Earth (Locke 1988 [1689], §50; emphasis added).¹

Consenting to money, for Locke, implies consenting to the distribution of benefits and burdens caused by monetization.² It follows that the consenting parties cannot cherry-pick some of the externalities of monetization out of the scope of their agreement. Nor can they renege on their voluntary commitment to the use of money if the forces of the monetary economy land them in poverty.

In this paper, I argue that the consent argument (and its traditional interpretations) is beset by two problems, which jeopardize its ability to deliver the sweeping justification of material inequalities that Locke and his followers thought could be derived from it. Regarding the consequences that the consent to money is taken to legitimate, contemporary neo-Lockeans³—and, to the extent that he held this view, Locke himself—are wrong in holding that the scope of justified inequality that is warranted by the device of consent also covers inequalities in the original appropriation of land. In contrast, I argue that the consent argument can only deliver a justification of the inequalities resulting directly from the monetization of commercial exchange. In this respect, defending inequalities in original acquisition by appeal to the idea of consent to money constitutes a conflation of the (independent) categories of justice in transfer and justice in appropriation.

¹ Locke 1988 [1689]. Hereafter, numbers in brackets after '§' refer to the paragraph of the *Second treatise of government*.

² Whether consent to money is to be thought of as actual or merely hypothetical is an independent problem which I shall not discuss in this paper.

³ It is difficult to give a synthetic definition of what makes a political theorist a neo-Lockean. For the purpose of this article, I wish to include under this label not only the critics and interpreters of Locke's writings, but also those whose political thinking is methodologically or substantively germane to Locke's own political philosophy, including (non-constructivist) contractarians, and (right- and left-) libertarians.

Regarding the *nature* of the consent to money, neo-Lockean critics perpetuate Locke’s own failure to adequately appreciate the political underpinnings of the institution of money. While some have questioned the proposition that monetization leads to normatively justifiable outcomes (notably, Tully 1980), few Lockean critics have ever interrogated the assumption that monetization is indeed economically possible in pre-political societies. Only a minority (e.g., the chartalists) has suggested that the institution of money has an irreducibly political character that makes its establishment and justification impossible to decouple from the establishment and justification of political authority (e.g., Bell, et al. 2004). In the course of my discussion, I develop the proposal that money is a quintessentially political creature. My argument seems to lead to the refutation of Locke’s own idea that money is a state-of-nature institution, and to the conclusion that monetization coincides with the constitution of a political society—or, more precisely, a “political economy”.

I conclude the article by exploring the consequences of my defense of the political nature of money in light of Locke’s argument that consent to money justifies the inequalities induced by monetization. The political interpretation of the consent to money has far-reaching implications as to how the neo-Lockean is to theorize trans-national inequalities and global monetary institutions. In particular, it will emerge that, when the political account of money is situated in global context, the range of inequalities that the consent argument is apt to justify is confined to economic disparities within political jurisdictions. Since consent is expressed through the political compact, trans-national inequalities cannot be said to be thereby justified.

Because of these misconceptions about the consequences and the nature of the consent to money, the programmatic justification of economic inequality that Locke deploys through his consent argument is vulnerable to two egalitarian challenges. The joint force of these challenges seems to compel neo-Lockeans to scale down the range of inequalities that can be thought to legitimately arise from the introduction of money. Inequalities in original appropriation and trans-national inequalities emerge as impossible to justify with reference to the consent to money. In the following section, I shall present my critique of Locke’s justification of acquisitive inequalities. The discussion of trans-national inequalities, which builds on my political

critique of Locke's account of the nature of money, is presented in the last section.

MONEY INEQUALITY: JUSTICE IN APPROPRIATION OR JUSTICE IN TRANSFER?

Money-induced inequalities and their justification

Locke's discussion of justice in transfer is almost entirely concerned with the economic inequalities that emerge in pre-political societies following the advent of money. It should be noted from the outset, however, that in Locke's view, money is not the only or the earliest driver of income polarization.⁴ Pre-monetary exchange is itself an important cause of material inequality. It is not difficult to see why. Imagine a pre-political society of producers-traders each endowed with a different level of labor productivity. It is reasonable to think that the more productive or able-bodied will enter the marketplace with a larger tradable stock, and thus greater bargaining power, than the less productive or disabled. Clearly, initial inequalities of tradable endowments, as aggravated by the bargaining advantages that the better endowed can gain in the market, result in income inequalities as the market clears.

Still, these inequalities do not seem to worry Locke. He claims that, as long as wastage is eschewed, "any one can make use of [the income from one's labor] to any advantage of life" (§31). And this permission must be taken to include not only the possibility of directly increasing consumption, but also the possibility of improving one's trading position and bargaining leverage in exchange. It follows that, as he puts it, "if [someone] bartered away Plumbs [sic] [...] for Nuts [...], he did no injury" (§46); not even when it turns out that, given the circumstances, swapping out plums for nuts benefits the seller of plums more than the buyer of nuts in relative terms. Locke would conclude—although he never states it explicitly—that relative inequalities of outcome are morally irrelevant, as long as exchange is consensual.

Justifying uncoerced trade in kind, and the inequalities it generates, is merely a starting-point. Economic agents engaged in barter would soon transition to a monetary economy by entering into a "tacit Agreement of Men to put a value" (§36) on unitary quantities of a

⁴ By "income" I mean any material advantage (whether in cash or kind) that can be derived by using or selling one's labor, or through commercial exchange.

designated durable substance (Locke 1991a [1691], SC.145),⁵ such as a “sparkling Pebble or a Diamond” (§46). Locke does not discuss at length why agents would choose to introduce a tradable currency, focusing instead on the question of how monetization is possible.⁶ Yet it is clear that, for Locke, the core function of money is to enable its holders to preserve from dissipation the value embedded in the “truly useful, but perishable, Supports of Life” (§47), namely consumable commodities of the likes of plums and nuts. Money gives producers and traders the capacity to accumulate economic value and avert the so-called “spoilage proviso”, i.e., the natural-law requirement that nothing be spoiled or destroyed (§31).

Despite its advantages in the way of efficiency, the monetization of trade raises normative questions in its own right, for it causes an additional, and much deeper, wave of income polarization than that occasioned by non-monetary commerce. Money-induced inequalities pose justificatory challenges that cannot be settled simply by appealing to a history of consensual commercial transfers, as in the case of inequalities in the pre-monetary phase. If we are to justify the deeper unequalizing forces that monetization precipitates, economic agents must consent to the very use of money as an instrument of exchange, and not just to each and every transaction concluded in cash.⁷

I will be concerned with this two-fold question: what exactly are the distributive consequences of monetization; and which of these consequences can be genuinely thought to be legitimated by reference to an act of consent? In tackling this problem, most commentators, and more controversially, Locke himself, seem to make a conspicuous mistake, which leads them to conflate the realm of justice in transfer with that of justice in appropriation. Let me first reconstruct in some detail the arguments found in the literature. The view I contest hinges on Locke’s assertion that where money is introduced, “Men will [...] be apt to enlarge their *Possessions of Land*” (§48). The possibility of accumulating economic value through money leads economic agents

⁵ Locke 1991a [1691]. All citations from this work are marked as ‘SC’, and refer to the paragraph number as labeled in this edition.

⁶ See Caffentzis 1989, 73. I discuss this problem in the next section.

⁷ Of course, one could reasonably question whether consenting to the institution of money is, though necessary, also sufficient for a compelling justification of money-induced inequalities. One could argue, for instance, that the ramifications of the introduction of money are so extensive and pervasive that consent to the institution of money cannot signify consent to the material consequences of monetization. In what follows, I shall not take up this challenge and will instead assume that the core thrust of Locke’s consent-based argument is sound.

to appropriate more land than they could "use the product of" (§50), in order to sell the "overplus" at a price (§50) and accumulate currency. Thus, money "made Land scarce" (§45); and as resource scarcity inevitably curtails the acquisitive opportunities of some, all critics seem to converge on the view that the introduction of money is (or, at least, appears to be) at loggerheads with Locke's "sufficiency proviso", that is, the natural-law requirement that original acquisition should leave "enough and as good" resources for others to appropriate (§27, §33).

Money and the law of nature

How do neo-Lockeans reconcile the egalitarianism of the "sufficiency proviso" with the inequalities allegedly permitted by positive consent? There are two ways of resolving this tension. Accordingly, neo-Lockean critics can be sorted into two camps. Those that I shall call "abrogationists" argue that, because monetary economies fall foul of the sufficiency proviso, common consent is necessary to repeal or abrogate the Law of Nature and make monetization possible (Ince 2011, 36-37; Waldron 1988, 220; Macpherson 1972, 211). Adopting a more critical stance, other abrogationists view the consent to money more as a blatant "violation" than a permissible "abrogation" of the Law of Nature (Ince 2011, 37). Because of its seditious character others argue money not only violates the norms of the pre-monetary order, but also causes the pre-political economy to become dysfunctional; and the social instability that ensues provide powerful motives for the abandonment of the state of nature and the establishment of civil rule (Tully 1980).⁸

Instead of attempting to transcend the Law of Nature through the mechanism of consent, the "*revisionists*" choose to strategically re-theorize, or revise, the Law itself. They maintain that although the "land grab" triggered by monetization leaves some without enough and as good in the way of natural resources, the benefits that accrue to the propertyless by way of new opportunities for employment, rental and purchase outweigh the opportunity costs of missed acquisition (Mack 2009, 70; Sreenivasan 1995, 35-37). The proviso must be simply satisfied *all things considered* and the consequences of monetization do

⁸ Tully (1980, 154) makes this point very eloquently: "Money disrupts [the] natural order, and government is required to constitute a new order to social relations which will bring the actions of men once again in line with God's intentions". See also Caffentzis 1989, 68.

not seem to be in breach of this more relaxed standard.⁹ A corollary of this is that, because the satisfaction of the proviso is sufficient to legitimate the consequences of the introduction of money, common consent plays at best an ancillary role in the justification of money-induced inequalities (Mack 2009, 67).

What is the common ground shared by these two interpretive strands? Both views are premised on Locke's thesis that monetization can justifiably lead to a “land grab” (whether in compliance with the Law or by overriding agreement) and, ultimately, to resource scarcity. However, I think there are reasons to question the adequacy of Locke's avowed view of the consequences of monetization, based on both textual and analytical considerations.

Money, appropriation and innovation

For one thing, there is countervailing textual evidence suggesting that, for Locke, what is actually incentivized by the introduction of money is not further land appropriation, but simply labor and industriousness. In principle, a single appropriator could acquire extensive land holdings, bring them into cultivation and, by selling the product she does not need for her subsistence, hoard up currency. However, Locke seems to concede that the “part of [the original commons that] the industry of one man could extend itself [to]” is in practice very small (§31, §36), even when opportunities for permissible accumulation are opened by the institution of money. To be sure, our ambitious appropriator could circumvent this constraint by renting out or selling the newly acquired surplus land.¹⁰ However, the *Second treatise of government* contains no explicit mention of commercial transaction in land;¹¹ and the primary source of accumulation in the state of nature is supposed to be production-driven trade: the “larger Possessions” (§36) introduced by money are made possible by the sale of the likes of nuts, sheep and wool (§46)—that is, consumer goods—rather than by the sale or rent of land.

⁹ This argument may have been implied by Locke when he says that one can legitimately accumulate money not only because consent to the resulting inequalities is implied in the common consent to money, but also because hoarding up gold and silver occurs “without injury to anyone” (§50).

¹⁰ In fact, Locke's economic theory does, consistently with this hypothesis, include the notion that land is capital, i.e., as a good capable of “yielding a certain yearly Income” or rent (SC.25).

¹¹ In keeping with this, Tully (1980, 149) maintains that Locke did not think of land as capital, or as a capital good.

Given the organizational challenges that—as Locke acknowledges (§31)—are associated with bringing new land into productive use (and absent appropriation for rent-seeking purposes), it is reasonable to conclude that Locke occasionally thought that what the institution of money really generates is an incentive to innovate and boost efficiency in the productive use of already acquired land. Also other passages seem to contradict the mainstream view that monetization leads to a “land grab”. For instance, Locke suggests that the reason why the (non-monetary) societies of America “have not one-hundredth part of the conveniences [England] enjoy[s]” is not so much that the former have yet to bring unimproved land into cultivation as the lack of incentives for “improving [the land] by labour” (§41). The most conspicuous difference between the incentive structure of monetary and non-monetary economies is that the former promotes labor, industriousness and innovation, whereas the latter merely ensures subsistence.

Moreover, the result that monetization incentivizes innovation rather than appropriation seems to follow from Locke’s idea that (in situations of non-scarcity such as the state of nature) economic value is determined almost entirely by labor, rather than as a result of physical inputs such as land (§40, §42).¹² If land is all but valueless, the possibility of storing value by hoarding currency must be taken to offer incentives to deploy ever more complex forms of labor,¹³ rather than to enclose unimproved land.

In sum, several passages in Locke’s *Second treatise of government* suggest that the primary incentive provided by money is to promote innovation and industriousness, rather than further appropriation. The crucial implication of this claim is that the material inequalities that market participants consent to as they consent to the institution of money are those that result from the differential capacities of market participants to engage in innovation and industry. Money cannot possibly enable market participants to realize benefits from further appropriation. Consequently, consent to money cannot be taken to imply consent to the sanctioning of such (unrealizable) benefits.

¹² In situations of resource scarcity, Locke instead espouses a demand-supply theory of exchange value (see §45; and SC.170).

¹³ See the passage (§43) where Locke lists the many types of value-adding labor that go into making the final value of consumption goods.

The justifiability of appropriation

Textual considerations aside, there is a second (this time, normative) reason why Locke’s consent argument should not be seen as justifying inequalities in original appropriation, whatever Locke’s positive pronouncements on this issue (§49, §50) might have been. Even if state-of-nature landowners had, in a conjectural history, attempted to rely on the appropriation of marginal land to kick-start capitalist accumulation, the normative clout of Locke’s Law of Nature would simply have barred this course of action. Both revisionist and abrogationist strategies to resolve the (real or apparent) clash between the sufficiency proviso and the consequences of monetization fail to get off the ground. Let me elaborate with reference to two of Locke’s famous formulae. Firstly, Locke states that labor is “the measure of Property” (§36). Undoubtedly, this formula means, positively, that labor gives rise to property relations (i.e., the labor-mixing argument). But it also prescribes, negatively, that one may not appropriate more land than one could either labor on by oneself, or (more controversially) hire manpower to labor on (§28). In other words, property is (at least primarily) to be used for productive investment, rather than rent-seeking. Consider now the second formula: Locke says that each instance of appropriation must leave “enough, and as good [...] *in common* for others” (§27; emphasis added). This proviso (i.e., the “sufficiency proviso”) means, negatively, that, even if somebody was such a “production monster” that he could labor (or invest) on vast tracts of land,¹⁴ the appropriation thereof would be barred. The proviso also means, positively, that everybody is entitled to an opportunity to expend one’s labor, and hence enter into property relations.

Recall now that the revisionist argues that non-appropriators are compensated for remaining propertyless by means of opportunities to labor on the property of others, or to purchase manufactured goods. But it is clear that the prescriptions that these formulae contain are much more demanding than those implied by the revisionist proviso.¹⁵ What is to be left for others to acquire, according to the second formula,

¹⁴ This would satisfy the requirement that labor be a “measure of Property”. Let us also assume that, besides a “production monster”, such appropriator is also a “consumption monster”, so that nothing is wasted and the “spoilage proviso” (§31) is satisfied. This example is modeled on Nozick’s example of the “utility monster” (1974, 41).

¹⁵ I am not arguing here that it would be wrong to endorse, on independent grounds, a (revisionist) proviso that constrains, not the distribution of resource inputs strictly speaking, but the overall distribution of final utility levels. All I am saying is that Locke’s own wording of the proviso is more akin to the former type than to the latter.

is not just a stock of *produced* resources, but a stock of *productive* resources as found in the original commons. Furthermore, in light of the first formula, acquisitions of marginal land solely for rent-seeking purposes unacceptably eviscerate the function of property as the vehicle of productive investment.

It is also unsatisfactory to reply to this argument, as the abrogationist does, that the proviso interpreted strictly is abrogated by the act of consent to money. For the same reasons that mutual or informed consent are not sufficient to legitimate slavery or suicide—they are impermissible under the Law of Nature—by the same token, Locke must think that positive consent cannot override the natural Law more generally (Kelly 2007, 106; Vaughn 1980, 92-93). In fact, Locke's contention that rightful ownership does not depend on the consent of the rest of mankind (§28) makes positive consent not only insufficient but also unnecessary for the justification of unequal appropriation.

Revisionism and abrogationism incur the same charge. In thinking that monetization of exchange leads to acquisitive inequalities, both interpretations of the consent to money conflate the category of justice in transfer with that of justice in appropriation: the normative procedure (i.e., consent) that is supposed to regulate inequalities arising directly from monetized trade is read back into the normative sphere of appropriation, and wrongly taken to warrant original acquisition of land beyond the limits set by the norms of just appropriation (as codified by the two formulae discussed above). Consent to the use of money only entails consent to the *consequences* of the use of currency as a mechanism of resource transfer. It surely cannot legitimate *all* consequences, including the putatively greater incentives facing currency traders to engross their land stocks. If so, neo-Lockeans would also run into further paradoxical conclusions, namely that consent to money would legitimize the crimes produced by the "state of contention [and] covetousness" afflicting monetary societies (Tully 1980, 150), not to mention the crimes of coin-clippers and counterfeiters (Caffentzis 1989). In sum, contra the revisionist, a correct understanding of Locke's Law of Nature would strike down the presumption that the inequalities occasioned by a money-driven "land grab" can be justified on Lockean grounds. Furthermore, Locke's Law of Nature cannot be repealed by positive consent.

What should neo-Lockeans make of the passages in the *Second treatise of government* that suggest the consent to money implies the

consent to unequal possessions “of Land” (§48, §50)? In light of the textual evidence reviewed above and the analytical tensions raised by Locke’s avowed extension of the consent argument into the realm of appropriation, it is reasonable to conclude that these passages reveal, at best, an internal contradiction within Locke’s own account of money. Locke’s low opinion of mercantile activities and his belief that “agriculture was the foundation of English society” (Wood 1984, 20) might go some way towards explaining his assertions that the consent to money legitimizes inequalities in land holdings, and his willingness to discount the tensions between these assertions and his other pronouncements on money and natural law. Absent a convincing account that brings paragraphs §48 and §50 to cohere with Locke’s provisos, the need to preserve analytical consistency within Locke’s normative architecture imposes that the weight of these passages—and the orthodox interpretations of Locke’s consent argument drawn from them—should be considerably qualified.

Trade-related inequalities

Let me clarify. I am not arguing that the introduction of money would not generate inequalities, or that consent legitimates none of the inequalities found in monetary economies. Rather, I have tried to show that, for Locke, inequalities in the original appropriation of land cannot be set right by consent. Importantly, the class of inequalities that result directly from monetized exchange remains in need of justification and their legitimacy cannot but rest on the sanction given by common consent. In order to justify the much more pervasive inequalities found in monetary economies, Locke must posit that their institutional determinants—i.e., the adoption of a common means of exchange—should be consented to as such. Even when it is granted that it cannot vouch for acquisitive inequalities, the consent device is far from vacuous. The revisionist is thus mistaken in maintaining that consent performs a merely ancillary function in Locke’s justification of economic inequalities.

The revisionist could retort that since accumulation is made possible by the greater labor productivity of the “Industrious and Rational” (§34), and “the *Labour* of [one’s] Body [...] [is] properly [one’s own]” (§27), the inequalities from monetization can simply be justified as consequences of the self-ownership thesis. However, accumulation is socio-economically possible owing to the availability of money as a

value-storage institution and as a commonly recognized means of exchange. The industrious and rational could not reap the material benefits of his natural endowments in the absence of a monetary system. Therefore, the premium captured by the able-bodied requires independent moral justification. Locke's idea of consent to the institution of money precisely delivers such justification, thus qualifying as a more than a decorative device.

Yet it would also be wrong to overestimate the range of inequalities that the consent device is apt to legitimize, as some have done by claiming that "there is no limit, apart from the amount of [...] [productive] labor [one] commands, on the extent of [accumulation]" (Waldron 1988, 220; Macpherson 1972, 204). Locke's spoilage proviso entails that a producer is not allowed to produce more than the market could possibly absorb (Weymark 1980). Thus, aggregate consumptive capacity represents the limit that Locke's Law of Nature imposes on the extent of rightful accumulation. In sum, the growth of inequalities from the level induced by unequal talents to the level vouched for by the aggregate spoilage limit represents the window of maximum unequalization that common consent to money is designed to legitimate. Justifying either land grabs or wastage sprees falls beyond the scope of Locke's consent device.

CONSENT TO MONEY: POLITICAL OR PRE-POLITICAL?

Tacit and expressed consent

If the economic forces set in motion by the monetization of trade produce inequalities, then consent continues to perform an important justificatory function. But how are we to think of the act of consent to money? In this section, I move from a discussion of the consequences of consent to the analysis of its nature. I shall return to the implications of my account of the nature of money for the justification of economic inequalities in the following section. The neo-Lockean literature has systematically failed to critically challenge Locke's understanding of consent to money, which—I submit—suffers from several flaws. In particular, I shall suggest that what is theorized by Locke as an act of *tacit* and *pre-political* consent is, upon closer inspection, an instance of expressed and political consent. This two-pronged claim leads to the conclusion that the monetized economies are irreducibly political and cannot develop in the state of nature.

Both revisionists and abrogationists hold that monetization develops in a pre-political economy, and “without Compact” (§50)—i.e., without political consent. In seeing money as the outcome of either a pre-political convention or of natural markets, they effectively theorize money as an institution of a supposedly pre-political economy. Both these interpretive approaches fail to capture the political character of the institution of money. As a result, they perpetuate Locke’s mistaken view that the emergence of monetary economies in the state of nature is both economically and normatively possible. To be sure, Locke’s pre-political view of money is certainly substantiated by the textual evidence: consent to money is a “tacit Agreement” that occurs “out of the bounds of [political] Societie” [sic] (§50; emphasis added). Thus, my argument cannot amount to an interpretive objection. Rather, it purports to be a critique of Locke’s argument itself and a denunciation of the uncritical endorsement of this argument that pervades the neo-Lockean literature.

My contention naturally invites an analogy with Locke’s argument that property-holders commit themselves to political obligations by tacitly submitting to the coercion and protection of government: “every Man, that hath any Possession, or Enjoyment, of any part of the Dominions of any Government, doth thereby give his *tacit Consent*, and is as far forth obliged to Obedience to the Laws of the Government” (§119). Tacit consent is delivered by silently adopting (and benefitting from) a given convention, in this case the power of governments to safeguard property interests. A similar argument can be run in the case of money: any economic agent that accepts any quantity of the designated value-bearing currency in a given marketplace, does thereby give her tacit consent to the social use of money in that marketplace, thus binding herself to not contesting the distributional consequences of such use.

Yet Locke’s main strategy to ground political obligations is a different one. He says that civil societies are constituted by expressed consent, with men “agreeing with other Men to join and unite into a Community” and thus becoming “perfect Member[s] of that Society, [and] Subject[s] of that Government” (§95, §119; emphasis added). Again, a similar line of argument can be developed in the case of money: agents explicitly agree with other agents to form a monetary convention within a given marketplace, and thus become perfect members of the convention and subject to the obligations imposed by the use of money.

Now, while Locke seems to think that, from a state-of-nature perspective, tacit consent is the mechanism that legitimizes monetization, he holds that it is through expressed consent that commonwealths are originally constituted (§122).¹⁶

Money and political obligations

The first question I want to discuss is to what extent monetization is, generally speaking, analogous to the creation of political obligations. In order to do this, I need to first elaborate Locke's conception of money. While other commodities carry value because they are demanded for consumption (SC.52), Locke thinks that money carries value as a “Pledge to procure” consumable goods (SC.31; see also Appleby 1976, 55). This formula encapsulates two ideas. First, a unit of currency is a unit of value that can be, generally speaking, retired through exchange rather than consumption. Second, money also carries a *promise* (a “Pledge”) to pay the value which is retired through exchange. And currency-holders can level this promise against all other parties to the marketplace. Note that the definition of money as a token embodying a promise to pay is distinct from the question of what substance can effectively carry this promise—such as the ounce of gold (SC.31), or—more generally—“any lasting thing that Men might keep without spoiling” (§47), or a paper banknote.¹⁷ However, we shall see that defining money as a “Pledge to procure” carries important implications for how we should think of the origins of money, as well as for what physical medium of exchange can genuinely count as money.

As a promissory token, a unit of currency is essentially a contract. Moreover, unlike other securities such as debt-bonds, currency encapsulates a contract that imposes rather burdensome duties on market participants. For it obliges to pay the currency-holder the agreed-upon value, not at a fixed future point in time (i.e., at maturity), but on the currency-holder's demand; and the obligation falls not just on a specified “issuer”, but on any market participant (SC.32). Therefore, quite like the political obligations imposed by the social contract, the

¹⁶ However, property-owning foreigners (or members of later generations) consent to assuming political obligations only tacitly (§121).

¹⁷ Admittedly, however, in the *Second treatise of government* (§47), Locke seems to *define* money as “any lasting thing” (I thank an anonymous referee for noting this point). However, as I suggest later on in this section, this definition (as against Locke's definition of money as a “Pledge to procure”, which he gives in his economic writings) reflects a confusion between money (the “Pledge”) and its physical *body* (the “thing” that embodies the “Pledge”).

payment obligations imposed by the monetary contract are temporally continuous and equally affecting all parties to the monetary convention.

This seems to demonstrate that the monetary and social contracts bear important analogies. But we have not certainly succeeded in showing that monetization is a product of the social contract—that is, that monetary economies are inherently political and thus cannot precede the formation of political society. In order to arrive at this conclusion, I have to demonstrate (contra Locke) that monetization becomes economically possible (let alone normatively justified) only when market participants deliver common expressed consent to it (call this “express-consent thesis”). Further, I have to show that the procedure of expressly consenting to the institution of money (i.e., the monetary contract) is tantamount to concluding a social contract (call this “political-consent thesis”). Together, the “express-consent thesis” and the “political-consent thesis” entail the thesis that money is a political institution, and that monetary economies cannot develop in the state of nature.

Contracts in the state of nature

With a view to defending the “express-consent thesis”, let me first review a number of contracts that, according to Locke, economic agents can enter into in the state of nature. Just like money and political obligations, these contracts can be classified along two dimensions: the temporal distribution of the obligations they impose, and the number of parties involved. In the state of nature, a contract involving many parties and imposing obligations to be discharged within an extended timeline is to be regarded—I submit—as a paradigmatically risky contract. Conversely, a contract involving few parties and imposing immediately dischargeable obligations is to be regarded as a paradigmatically risk-free contract. This is because the future benefits that a party is contractually entitled to will be heavily discounted in the absence of an enforcing authority providing assurances that future obligations will effectively be honored. Moreover, ensuring compliance from many duty-bearers is, generally speaking, more costly than ensuring compliance from few.¹⁸

¹⁸ The social contract would itself be a very risky one if it did not include provisions to finance an enforcement authority, so that the costs associated with risk are mitigated, monetized and distributed amongst the parties to the contract.

The most elementary contract that Locke discusses is barter (§46). A contract regulating the exchange of, say, plums for nuts is, on our taxonomy, rather risk-free. It is temporally limited, as the discharge of transfer obligations by the parties occurs more or less simultaneously; and it is bilateral, so that compliance can be managed at relatively low costs. A second, more complex contract implied by Locke is wage-labor (§28).¹⁹ This is another bilateral contract (between worker and capitalist) but, unlike trade, the obligations of each party (to provide their labor, and to pay a wage, respectively) are dispersed over time; and time-dispersion introduces a dimension of risk. Insofar as we can regard original appropriation as a contract—where non-appropriators tacitly consent to abiding by non-trespass duties²⁰—appropriation would be an example of a time-dispersed *and* multilateral contract (that is, a risky contract). However, the risks associated with becoming a party to a “property contract”—that is, the risks associated with appropriating land and investing one’s labor on the understanding that the contract will be complied with—are somewhat mitigated by the fact that non-appropriators bear merely negative duties (to not interfere with the appropriator’s property). And, in general, we can assume that it is less costly to ensure that a (large) set of agents do not do certain things (e.g., knock down my fence), than to instruct them to actively do certain other things (e.g., pay me a sum on demand).

Consider now the monetary contract. As I noted earlier, the duties imposed by monetization are both time-dispersed and directed at a manifold of market participants. Furthermore, unlike in the “property contract”, the participants that agree (tacitly, for Locke) to be parties to the monetary contract are contractually obliged to discharge a positive duty whenever a currency-bearer presents them with currency. This duty consists of paying the bearer the (agreed-upon) value associated with the amount of currency he turns up. And paying the value will demand transferring to the currency-bearer a bundle of consumable commodities carrying that value. Of course, the obligation to pay the currency-bearer does not entail that currency-bearers may legitimately force sellers to sell their goods against their will. The only implication of

¹⁹ Whether Locke does believe that wage-labor relations are possible in the state of nature, and whether they are consistent with the Law of Nature, are highly controversial issues (see Tully 1980, 135-139; Waldron 1988, 225-232; Mack 2009, 60). I wish to remain neutral vis-à-vis this debate.

²⁰ This is arguably an overstatement, as Locke thinks that the original common can be particularized without the “express Compact of all the Commoners” (§25).

this “pledge to procure” is that a trader may not opt out of an otherwise consensual transaction merely on the grounds that it is denominated in a given currency. As it were, traders may refuse to sell goods, but may not refuse to buy currency. On all counts, the monetary contract is far from risk-free.

The “express-consent thesis”

Now, with the purpose of moving from the proposition that the monetary contract is not risk-free to the conclusion that its implementation requires expressed consent, let me consider an objection. My understanding of the monetary contract as a multilateral agreement could be contested. Waldron has argued that it is not necessary that consent to the value of money should be given by all market participants; all is necessary is that “those who are going to be parties to [individual] monetary transactions agree” (1988, 223-224). In explicitly consenting to the transaction, the parties would tacitly consent to the instrument’s value. The problem with this view is that, if monetary instruments carry value only for the two parties engaged in a transaction, there is no reason to think that the two parties would enter into the transaction in the first place. After all, the point of accepting a commodity carrying solely exchange- (and no use-) value²¹ was precisely that the value embedded in the commodity could be accumulated and later retired through exchange with other market participants. As Marx rightly observed (1977, 443), monetary transactions are not purely private (like consumption-oriented barter), but bear a distinctly social character. However, if nobody but me recognizes the existence of value in the currency I am trying to sell, then it is irrational for me to even accept the currency in the first place. The risk of incurring a loss (by being stuck with currency that can neither be consumed nor traded for consumables) would be too high.

Two observations can be made. First, since demand for exchange goods depends, as Nozick rightly observes, on their “initial independent value” (1974, 18), and the value of currency is established “by Agreement” (§46), it follows that a currency market can only develop after a (multilateral) monetary agreement. Second, since we have seen that the possibility of entering into a money-denominated transaction—

²¹ Of course, gold carries use-value in the manufacturing industry, and as the supplier of the “body” of metallic money (Marx 1977, 444). But here the commodity I am referring to is not gold as such, but *money* itself, whose sole use is to enable exchange.

thereby delivering tacit consent to the value of the currency transacted—itself presupposes a prior monetary contract, it follows that such prior monetary contract can only be implemented by expressed consent. In other words, since the monetary contract could not itself be tacitly embedded in monetary transactions, it must, by exclusion, be concluded by prior expressed agreement.²²

It is not true, then—as alleged by Nozick—that “no express agreement and no social contract fixing a medium of exchange is necessary” (1974, 18): nobody would enter a market where valueless goods are traded. The benefits of monetization are only captured through “co-ordination economies”, which exist when certain market decisions are conditional on the decisions of other market participants. When all agents take coordinated decisions—say, to accept currency in market transactions—monetization increases overall efficiency by cutting transaction costs and facilitating the allocation of the “truly useful [...] Supports of Life” (§47). Yet such advantageous decisions would not be taken if there was uncertainty over whether or not all other agents (or at least a critical mass thereof) would indeed take the complementary decisions necessary for the expected advantages to materialize. Traders would arguably demand more tangible assurances to mitigate the uncertainty and perceived risks of monetary transactions. So, monetization could only get under way after a procedure of express common consent has been concluded:²³ entering monetary transactions would otherwise be at best risky and at worst irrational.

The “political-consent thesis”

This seems to prove what I called the “expressed-consent thesis”. However, we could infer that monetization cannot precede the social

²² This argument is predicated upon the neoclassical view that economic value is created in exchange (rather than in production). However, it adds to this view by suggesting that, in the case of money, value can emerge through exchange only provided a background monetary contract is established. Effectively, this argument charts a statist view of money. If alternatively we begin with the (Marxist) view that economic value is created in production, we would arrive at the view that (commodity) money can acquire value even in the absence of a monetary contract. See, for instance, Lapavistas 2000.

²³ The thesis that the introduction of money requires a prior generalized contract is obviously amenable to being tested against historical or anthropological evidence (I thank an anonymous referee for suggesting this point). In this paper, I only adduce *theoretical* arguments against the possibility of pre-political monetary systems. For some empirical evidence, see Bell, et al. 2004.

contract only if we could establish that the monetary and social contracts are concomitant or coincident; in other words, that the conclusion of the monetary contract marks the constitution of a political society (“political-consent thesis”). Otherwise, if expressed consent to the monetary contract was non-political, and could thus obtain in the state of nature, it could still be possible for monetization to antedate the constitution of political society. The point here is that consent to a common currency is a fundamental aspect of a political compact. Constituting the body politic means (amongst other things) constituting a *political* economy—that is, an economy that is organized through, and ruled by, political power. In this sense, a market that adopts a monetary instrument (by common expressed consent) has thereby constituted itself as a body politic. In entering the monetary contract, the parties to the marketplace organize themselves politically through the common pronouncement of consent to a generalized contract (i.e., the monetary contract)—which applies to and regulates all economic transactions amongst them.²⁴

Let me add a further point. Risk is associated not only with each monetary transaction effected in the absence of a prior monetary convention, but also with the monetary convention itself, whenever the latter is concluded without a common enforcing authority. Parties to the marketplace may subscribe a common “Pledge to Procure” (SC.31). But who guarantees that traders will actually “procure”? Thus, not only does monetization depend, for its economic viability (and normative justification), on the constitution of a body politic. It also requires the establishment of a civil government (call this the “political-authority thesis”). Political institutions are needed to compel market participants to accept currency in payment for consumables, in lieu of other consumables,²⁵ and to adjudicate disputes under the monetary contract.

²⁴ I agree with Nozick, though, that a marketplace, *as such*, “needn’t become a marketplace by everyone’s expressly agreeing to deal there” (1974, 18). While money is a political institution, (pre-monetary) markets have a natural and pre-political character—although their full development might require the intervention of political institutions (see Chang 2002, 547). Although seemingly counterintuitive, the different status of non-monetary and monetary markets is not contradictory. In fact, it is justified by the peculiar social features of money.

²⁵ Another function of the enforcing authority would be to prevent counterfeiting—a topic that Locke particularly exercised (see SC.146; see also Caffentzis 1989, 71).

Some objections

Let me now recap my arguments so far, draw out some of its implications and address some objections. We have seen that money is not just, as some critics have held, "the generating cause of the social contract" (Caffentzis 1989, 71; Tully 1980), but a social contract itself; and monetization is a feature of political rather than "natural" economies, so that it cannot be taken to be economically viable in a pre-political state of nature. With this picture at hand, we can now state another reason why abrogationist and revisionist neo-Lockeans are wrong in following Locke and seeing money as (respectively) a "conventional-historical" practice or a feature of supposedly "natural" markets: not only do such views distort what count as the permissible material consequences of monetization, but they both mistakenly de-politicize the institution of money.

At this point, the advocate of the view that state-of-nature markets spontaneously produce money could object that the political view I have defended fails to explain a number of cases where monetization seems to occur without either express consent or any centralized enforcement mechanism. A typical example is the spontaneous emergence of cigarettes as mediums of exchange amongst prisoners or soldiers at war. Let me briefly reply to this objection.

To start with, it is not clear that cigarettes are actually money as opposed to simply being the commodity most in demand (Ingham 2004, 24). After all, cigarettes are consumables and their role in exchange is explained by the fact that they are ultimately demanded for consumption. It is hard to imagine that a prison ward or battalion populated solely by non-smokers would adopt cigarettes as a medium of exchange. Thus, since the value of cigarettes in commerce is ultimately derived from their intrinsic properties, there is simply no need to even call on an argument from tacit consent to explain the exchange value of cigarettes. Of course, the opponent of the political conception of money could insist that tacit consent is necessary, and mere demand for personal consumption is not sufficient, to explain why cigarettes come to be used in prison trade. But even when backed by tacit consent, it is hard to think of cigarettes as money, for the holder of such "currency" would have no assurance (other than the one derived from the knowledge of the consumption preferences of other inmates) that it will be accepted in future exchange. But genuine money, as a "Pledge to

procure” (SC.31), does offer this assurance.²⁶ Furthermore, even if we conceded that prison cigarettes were genuine “money”, we would be hard pressed to explain how a monetary system allegedly based on tacit consent could be extended beyond artificially constrained markets such as prisons. Where the behavior and preferences of market participants are not always fully known (as is the case in real economies), monetization requires the expressed consent of all, as well as the support of a politically sanctioned enforcement system.

Lockean arguments for the political view of money

While my argument in favor of the political nature of money should be taken as an objection to, rather than an interpretation of, Locke’s monetary theory, it is important to stress that thinking of the monetary contract in political terms chimes well with some of Locke’s own propositions. For one thing, his main worry with state-issued paper bills was the risk of counterfeiting (SC.31). On the face of it, this concern seems motivated by the lack of anti-counterfeiting technology in Locke’s times. This made it (circumstantially) impossible to print bills that would be as difficult to replicate as the precious metal was to mine. But if Locke endorses metallism purely on circumstantial grounds, he would have to concede that, were paper bills made difficult to counterfeit, there is no reason why the law could not genuinely annex value to paper money. In an apparent recognition of the weakness of Locke’s rejection of paper bills, Geoffrey Ingham calls Locke a “practical metallist” (2004, 40). Locke’s objection to paper bills does not, by itself, entail a conclusive rejection of political consent.

More importantly, Locke’s treatment of coinage as a post-state-of-nature institution points in the direction of my argument. “The government of Politick Societies, introduced Coinage, as a [...] *Warranty* of the public” (SC.146); and “the Coining of Silver, or *making Money of it*, is the ascertaining of its *quantity* by a publick mark”—the mark being “a publick voucher that a piece of such a denomination [...] has so much Silver in it” (Locke 1991c, paragraph 5; emphasis added). While we cannot conclude that Locke thought that gold became *valuable* upon

²⁶ Even in cases of hyperinflation, while a unit of currency loses its purchasing power, it still does not lose its status as legal tender. Conversely, the mere economic *fact* that cigarettes are accepted in trade is not sufficient to turn cigarettes into *legal* tender, in the same way that Locke’s observation that the state of nature is a state of peace (§19) is not sufficient to turn the state of nature into the civil state. In both cases, there is a lingering risk that, respectively, the alleged currency (cigarettes) might not be accepted in trade and that peace may be destabilized.

coinage—as coinage is the mere “ascertaining” of already existing value—these passages suggest that political institutions play a central role in upholding the value of money. My argument is an extension of this point: since the consent to money is conveyed through the body politic, political institutions not only uphold but also create the monetary value of gold. Lastly, Locke’s attitude to monetary crime is also in line with the contention that some passages in Locke are not inconsistent with the idea of money as a political institution. Conspicuously, Locke says that coin-clipping and counterfeiting are “the highest Crime[s], and [have] the weight of Treason” (SC.146; emphasis added).

On balance, however, while there are grounds to relate my argument to Locke’s thinking, the political understanding of money I have derived in this section (starting from—it should be remarked—Locke’s own definition of money as a contractual “Pledge to procure”) is undoubtedly at loggerheads with the substantive content of Locke’s philosophy of money.

“Fancy” and metallism

In yet another punchy formula, Locke claims that value is assigned to currency by “Fancy or Agreement” (§46). Regarding the latter, we have seen in the course of this section that it is a mistake to think that agreement can be tacit, as Locke maintains. Regarding the former, I do not think, as Locke does, that traders in pre-political economies would accept gold by mere “Fancy”, as if “pleased with its colour” (§46). If my arguments in this section are sound, (expressed) “Agreement” is both necessary and sufficient to establish a monetary system. This makes “Fancy” redundant. Moreover, to think that parties to pre-monetary markets would accept gold in exchange for valuable commodities (the “Supports of Life”) is to attribute a fetishistic nature to the metal (Caffentzis 1989, 48 and 91). But, on the assumption that parties to pre-monetary markets are rational—rather than spell-bound by an irrational fetish—we must conclude that Gold would not be *per se* “fancied”. There is no market demand for gold in pre-monetary economies, for demand for a substance that has no “real Use” in supporting life (§46) can only arise after value has been assigned to it by fiat. As Richard Temple—a contemporary of Locke and critic of his monetary theory and policy—wrote, it is “the mony [sic] of every Country, and not the Ounce of Silver, or the [putative] intrinsick value

[of metal], [that] is the Instrument and Measure of Commerce there” (quoted in Appleby 1976, 50).

The objection to the creation of money (and its value) by mere “Fancy” seems corroborated by some commentators, who argue that, for Locke, gold money is not valuable in nature *as a commodity*, for gold is simply not a consumable. After all, Locke’s is a theory of fiat money rather than a “commodity theory”. What really does the work of explaining value, in the economy of Locke’s treatment of money, is the idea of tacit agreement (Kelly 1991, 89; Ingham 2004, 40). While Locke’s conception of consent is fundamentally different from the one advanced in this paper, both conceptions share the view that the value of money is created essentially by consent. Despite the appearances of his argument from “Fancy”, Locke could not genuinely have meant to commodify the physical body of money, for this would have entailed that there can be market demand for the natural attributes of such body. But the natural attributes of gold (its color, sparkle, hardness, and so forth) have no real socio-economic use.²⁷ The only valuable attributes of gold are its socially constructed powers—namely the powers, assigned to it by common consent, to command useful commodities on the market. To put it with another 17th century critic of Locke, James Hodges: “Silver, considered as Money, hath, speaking properly, no real intrinsick value at all”, for “the whole value that is put upon Money by Mankind, speaking generally, is extrinsick to the Money” (quoted in Appleby 1976, 51).

Before I round off this section, let me note that a salient implication of my rejection of both “Fancy” and tacit “Agreement” is that Locke’s detraction of paper money loses its theoretical rationale. In his economic writings, Locke holds that only gold can act as a “Pledge to Procure”: “a law [however expressly consented to] cannot give to [paper] bills that intrinsic value which [...] consent has annexed to [...] Gold” (Locke 1991b [1668], 173). But only if market demand for gold is “naturalized”, or tacit consent acknowledged, can Locke justify his definitive rejection of non-metallic money. This is because if money is valuable by “Fancy” or tacit “Agreement”, then paper bills could never

²⁷ Of course, there is (natural, pre-political) demand for gold insofar as gold (or other precious metals) can be used in jewelry and for the making of handicrafts. However, a putative state-of-nature market for handicrafts could hardly sustain the demand levels necessary to explain the circulation of gold *as money*. Furthermore, even if demand levels were high enough to justify the use of gold as medium of exchange, we would still be hard pressed to clarify why gold is, rather than “money”, simply the commodity most in demand.

acquire value: while there can be fetishism for the metallic body of money, there can be no fetishism for purely abstract, “disembodied” money; and while state-of-nature traders would liquidate their assets against gold—tacitly consenting to its value—they would never liquidate their assets against “a paper portrait of William III” (Caffentzis 1989, 118)—so that there could be no tacit consent to otherwise valueless paper bills. However, since “Fancy” and tacit “Agreement” are shown to be unsatisfactory explanations of money, then the possibility of fiat money reemerges forcefully.

Note that the material substratum of metallic currency performs two functions: it “carries” the value; and/or it “represents” the value through a physical body whose mass is proportional to the magnitude of the denomination. Locke thinks that both functions (but especially the second) are necessary for money to be valuable. However, if it is true that what is sufficient (and necessary) for monetary value is authoritative expressed agreement, then both the “carrier” and the “physical representation” of value become irrelevant. A unit of currency is, in its essence, an abstract power rather than a substance: it reflects a “willingness to accept an equality between it and [a certain physical commodity] that is not *in* it” (Caffentzis 1989, 75; see also Locke 1991c [1696], paragraph 2).

It seems that the political interpretation of the consent to money I have defended is compatible with the view that money is essentially an abstract entity, which can be signified through a paper bill guaranteed by the state. In other words, my political account of money is consistent with chartalism (Bell, et al. 2004).

UNIVERSAL MONEY AND TRANS-NATIONAL INEQUALITY: A LOCKEAN PRIMER ON GLOBAL MONETARY JUSTICE

Domestic or global money?

Neo-Lockeans of all stripes seem to commit a two-pronged fallacy by at once overstretching the justificatory scope of the monetary contract and by perpetuating Locke’s failure to appreciate the political character of this contract. As mentioned in the introduction, the purpose of jointly discussing the consequences and the nature of money is that my political interpretation of the monetary contract has far-reaching implications as to how neo-Lockeans should theorize material inequalities and monetary institutions at the trans-national level. In particular, endorsing a political account of money further curtails the

range of inequalities that the consent argument is apt to justify once we situate our discussion of money in global context.

The political interpretation casts an ambiguity embedded in Locke's thinking (and largely overlooked in the secondary literature) under a different, more troubling light. Throughout his discussion of consent to money, Locke remains ambiguous as to whom he regards as the consenting subjects. On some occasions, he sees the monetary contract as receiving the “*universal Consent of Mankind*” (SC.32; emphasis added; see also §45; SC.31; Locke 1991b [1668], 173). On other occasions, Locke seems to think, in line with my argument, that the consenting subjects are the members of domestic political jurisdictions. For instance, he implies this when he argues that the “Riches and Treasures taken away” during an unjust war “have but a Phantastical imaginary value” for the aggressor country (§184). Locke thinks that since the aggressor does not recognize the value of the currency used by the victim country, she is under no obligation to return the assets seized in the course of the unjust war.²⁸ The best explanation of this claim is surely that, for Locke, aggressor and victim are parties to distinct monetary conventions, and that consent to the value of monetary assets is given within the political boundaries of each convention.

For Locke, is the institution of money domestic or global? I think that the political conception of consent that I have suggested in this paper brings out a deeper dilemma built into this unresolved ambiguity. Let me first remark that the problem of identifying the subjects of consent is really an important problem only for those who, like me, emphasize the political character of monetary systems. If we assumed, as Locke at times does, that money was valuable by “Fancy”—that is, as a commodity—then it would be natural to recognize the universality of the institution of money. For if currency is in demand in light of the natural properties of its physical body, it would be implausible to argue that demand changes across political boundaries. Insofar as Locke espouses a commodity theory, and hence a form of metallism, we must then maintain that he should be committed to the universality of money, and we must regard his dithering over this problem of scope as simply ill-conceived. The fact that for the metallist the problem of scope has an obvious solution also explains why discussions of this problem

²⁸ This interpretation of the passage is also given by Waldron 1988, 223. The argument is obviously contestable, as the stolen assets do have, for the victim, real value grounded, arguably, in labor.

are so few are far between in the secondary literature. On the other hand, a neo-Lockean that maintains, like I do, that money is valuable as a result of a political procedure of expressed consent would have to determine whether such procedure is to be thought of as domestic or global in scope.

Trans-national inequalities

In this section, I do not intend to give a conclusive solution to this problem. Rather, I want to consider the implications for the justification of inequality of the domestic and global approaches to political consent. Let us start with the scenario where political communities establish distinct monetary contracts. Note, to begin with, that this interpretation seems to approximate a description of the existing global monetary order, where currencies are issued by national governments and are not linked to an underlying gold standard. In a world of separate monetary conventions, traders in, say, Australia would not recognize the value assigned to an Indian Rupee by the citizens of India. To be sure, Australian traders still accept rupiahs in international commercial transactions.²⁹ While it could be argued that, in accepting Rupee-denominated payments, Australian traders tacitly consent to the rupiah note as a store of value, it is clear that they only do so because there is an already constituted 1.2b-member-strong monetary union, namely India. Had India been a small, isolated country, traders would not rationally have entered into transactions denominated in Indian currency, and thus could not even be said to tacitly consent to its value.³⁰ So, consistently with the conclusions arrived at in the last section, tacit consent to a certain currency is not a sufficient condition for the emergence of trade in that currency. At least a segment of a market engaged in trade must have previously expressly consented to the currency's value, with other market participants "free-riding" on the consent of the first segment. And this is how international Rupee-denominated trade should be thought of.³¹

²⁹ Yet, as a matter of fact, most global trade is still conducted in the world's reserve currency, namely the US dollar. It is a question that my political approach to consent will have to address whether consent to a global reserve currency is indeed expressed and 'political' in my sense, or whether it is merely tacit.

³⁰ It is unsurprising that Bhutanese Ngultrums and North Korean Wons are in very low demand, and are thus close to valueless outside of Bhutan and North Korea, respectively.

³¹ In a similar fashion, the fact that trans-national trade in Locke's times was effected (also) through (unminted) precious metals should not be seen as a counterexample to

We are now in a position to identify another caveat on Locke's consent-based justification of material inequality. If a procedure of expressed consent is necessary (and sufficient) to justify the inequalities resulting from monetized trade, then the absence of a cross-country consent mechanism (other than tacit acceptance of payments) immediately delegitimizes all inequalities between countries engaged in commerce. For instance, any divergence between Australia's and India's (average) income levels that might be occasioned by monetized trade between the two countries would not fall within the range of permissible unequalization under Locke's consent argument.

To make this point clearer, imagine that the allocation of natural resources between India and Australia were just as measured against Locke's “sufficiency proviso”. In other words, imagine that Australia's unilateral appropriation of its iron-ore did leave enough and as good for resource-poor India to *appropriate* (and not just to *purchase* from Australia). The claim here is that, even against this background of justice in appropriation, all inequalities that would follow from bilateral monetized exchange would be unjust. This is because it is not the case that both parties (if either) consented (politically) to the monetary instruments used in bilateral trade—whether that be Rupees, Australian Dollars (or US Dollars). Therefore, a neo-Lockean would have to conclude that the ensuing inequalities are likewise unconsented-to (from the perspective of at least one of the two trading partners), and hence unjust. If consent is political, and the scope of consent is taken to be domestic, then international inequalities from trade cannot be legitimized by appeal to Locke's consent device.

Note that the argument that domestic consent within two countries does not justify trans-national inequalities between the two countries is independent of whether bilateral trade actually obtains or not. One can easily picture the relative inequalities that might develop between two (largely) non-cooperating countries (say, China and Taiwan) as a result of differential levels of efficiency in their internal (monetary) economies. The argument here is that such relative inequalities could not be legitimated by reference to China and Taiwan consenting to their own

the political account of money. After all, the demand for (unminted) gold is kept high because specie in the trading countries is coined out of gold. Traders engaged in long-distance commerce would accept bullion on the understanding that it could later be minted into coin. Once again, the behavior of traders in the monetary economy—i.e., their willingness to accept a monetary means of exchange—depends on the presence of a political authority (i.e., the mint) upholding the value of the means of exchange.

independent monetary conventions. To be sure, domestic inequalities occasioned by internal trade remain justified, as they are vindicated by the consent that Chinese and Taiwanese citizens confer on their own internal monetary systems. But this justification does not extend beyond political and monetary jurisdictions. The consent procedure in China sanctions the monetary value of the Renminbi. Consequently, it only justifies the opportunities for accumulation and the pressures toward unequalization occasioned by the social use of Renminbis. Even against a background of just resource appropriation, the relative benefits accruing to the Taiwanese by their consenting to the social use of Taiwan Dollars, instead of Renminbis, remain unaccounted for from the perspective of the Chinese.

The fact that Locke's consent argument is not applicable across the jurisdictions where consent is pronounced is no trivial matter. Note the relevance of the global economy that the current system where national currencies can be traded under floating exchange-rates is a major source of price instability, budget imbalances and unequalization (Wade 2009, 551). Thus, the range of inequalities that my argument shows to be morally illegitimate might actually be quite broad.

Universal money and global monetary institutions

At this point, the neo-Lockean might hope to circumvent my "egalitarian" conclusions by denying that Locke actually thought that monetary institutions are confined to domestic jurisdictions. As noted by Caffentzis, "Locke saw in the universality of money [...] the driving logical and social force of his age" (Caffentzis 1989, 119). While this stance is corroborated by the bulk of textual evidence (and, as mentioned earlier, it is the most reasonable interpretive option in light of Locke's metallism), the global approach to political consent still opens up the second horn of the dilemma. This is because the political interpretation of the "universal Consent of Mankind" (SC.32) demands that, in the state of nature, all parties to a putative global marketplace should constitute a global monetary convention supported by the requisite institutions for monitoring, enforcement or else. But since the gap between existing international monetary institutions and the ideal institutions demanded by a global monetary convention is quite significant (and indeed it is unlikely to be narrowed in the near future), the upshot of espousing the global approach must be a condemnation of the prevailing global monetary regime as unjust.

Let us look at the extent and nature of the shortfall between the ideal convention and the real-world order in some greater detail. While the existing order is best modeled by the domestic interpretation of consent, a real-world counterpart to the monetary institutions demanded by the global approach to consent could be the US Federal Reserve, as the issuer of the key global reserve-currency (the US dollar). The problem with arguing that the demands of the global approach are met by the near-universal use of the US dollar is that, from a state-of-nature perspective, many economic agents could not be said to have rationally consented to a global monetary convention where the dominance of a single national reserve currency systematically disadvantages deficit-ridden countries, and increases the likelihood of financial crises.³²

Alternatively, could the institutional role of enforcer of the global monetary convention be played by the erstwhile Bretton-Woods-era International Monetary Fund (IMF)? The Bretton Woods system was defined by the pegging of national currencies to a common monetary unit (i.e., gold). Therefore, in its capacity of global lender of last resort and arbiter of exchange rates, the IMF of the Bretton Woods era could be thought of as fulfilling the role of sovereign in the global monetary convention. However, for one thing, the undemocratic governance system of the IMF, which does not accord each country (let alone each global citizen) equal weight in decision-making, makes it an unlikely channel for a putative global consent to a gold standard system.

For another thing, a gold standard system may not be enough to realize the institutional demands of the global approach to political consent. On my interpretation of Locke's conjectural history of monetization, the parties to the pre-monetary marketplace would think it rational to establish not just a pegging regime, but a fully-fledged (supra-national) monetary union. This would require the implementation of a global currency to be used for private international transactions—rather than just for settlements amongst national central banks. The currency would be issued by a global mint and would have to be regulated by a global central bank, perhaps under the supervision of the IMF.³³ But the implementation of a global currency is a far-off goal

³² Interestingly, a similar point, yet from a more philosophical perspective, is made by Waldron 1988, 224. For the economic benefits on monetary stability from the use of a single supra-national reserve currency, see Wade 2006; Wade 2009, 550.

³³ Admittedly, a single supra-national currency may also generate imbalances and disfavor deficit countries, as the recent euro crisis has poignantly shown. If so, then

in a global monetary system where national currencies are not even pegged to a common monetary unit, but are subject to a floating exchange-rate regime.

A dilemma

The neo-Lockean that accepts my political interpretation of the process of monetization is thus faced with an impervious dilemma. If the neo-Lockean asserts—in an attempt to bring trans-national inequalities into the fold of justified inequality—that consent ought to be expressed globally, then she must condemn the existing international order as unjust on the grounds that it does not incorporate the one institution that parties to pre-monetary economies would rationally choose to establish, namely a global monetary convention. If on the other hand the neo-Lockean asserts that consent is delivered within discrete domestic jurisdictions, then it follows that the avowed capacity of Locke's consent device to legitimate money-induced inequalities comes out toothless beyond the domestic jurisdictions where consent takes place. The neo-Lockean that recognizes the political character of the consent to money is thus compelled to either denounce existing international monetary institutions or condemn the cross-country inequalities resulting from trans-national trade. On the political account of the nature of money, either the distributional outcomes or the institutional structure of the existing international system must be sanctioned as unjust.

CONCLUSION

The *problématique* of justice in transfer, and in particular the problem of monetization, has been marginalized in the secondary literature on Locke's economic philosophy, which has been dominated by the traditional issues of justice in original appropriation. Perhaps as a result of the paucity of discussion in this area, neo-Lockeans have failed to interrogate whether Locke's consent argument can really deliver the full vindication of money-induced inequalities that it promises. In this paper, I have argued that on closer inspection the range of inequalities occasioned by the use of money can only legitimately extend to

the parties to the monetary convention might find it rational to stop short of establishing a (rigid) global monetary union and, in its stead, return to a Bretton Woods-like gold standard regime with (flexible) exchange rates set by policy. In this case, gold would be the substance constituted as global legal tender.

economic disparities arising out of commercial transfers within definite political jurisdictions.

In the final analysis, an adequate understanding of the limitations of Locke’s consent argument entails an ideological shift to the left for all those theorists that endorse Locke’s Law of Nature and believe in the core thrust of his consent-to-money argument. While money may legitimately benefit the more industrious and deepen inequalities of outcome, it cannot dent the egalitarianism of Locke’s theory of original appropriation—which holds out through the monetization of state-of-nature economies.³⁴ Furthermore, since consent must of necessity be expressed and supported by coercive institutions, money-induced inequalities can only be justified within the political jurisdictions where consent is taken to have been conveyed.

Let me conclude by suggesting a reason why the political, anti-naturalist account of consent I have advanced offers a fecund theoretical standpoint. It is apt to capture and criticize the astonishingly unequalizing forces that—fuelled by trade and currency movements—make the global economy an often stormy playing-field. It is no surprise that Locke tried—though, to a large extent, failed—to give a self-standing justification for the material inequalities triggered by the monetization of trade: he was well aware of the socially destabilizing dynamics that afflict the global economy, once money makes its historical appearance:

People, Riches, Trade, Power, change their Stations; flourishing mighty Cities come to ruin, and prove in time neglected desolate Corners, whilst other unfrequented places grow into populous Countries, fill’d with Wealth and Inhabitants (§157).

REFERENCES

- Appleby, Joyce O. 1976. Locke, liberalism and the natural law of money. *Past and Present*, 71: 43-69.
- Bell, Stephanie. A., John F. Henry, and L. Randall Wray. 2004. A chartalist critique of John Locke’s theory of property, accumulation, and money: or, is it moral to trade your nuts for gold? *Review of Social Economy*, 62: 51-65.
- Caffentzis, George C. 1989. *Clipped coins, abused words and civil government: John Locke’s philosophy of money*. New York: Autonomedia.

³⁴ Effectively, this observation brings my critique of Locke’s pronouncements on the institution of money close to certain left-libertarian theories of distributive justice (e.g., Steiner 1994).

- Chang, Ho-Joon. 2002. Breaking the mould: an institutionalist political economy alternative to the neo-liberal theory of the market and the state. *Cambridge Journal of Economics*, 26: 539-559.
- Ince, Onur U. 2011. Enclosing in God's name, accumulating for mankind: money, morality, and accumulation in John Locke's theory of property. *The Review of Politics*, 73: 29-54.
- Ingham, Geoffrey. 2004. *The nature of money*. Cambridge: Polity Press.
- Kelly, Paul. 2007. *Locke's Second treatise of government*. London: Continuum.
- Kelly, Patrick. 1991. General introduction: Locke on money. In *Locke on money*, vol. 1., ed. Patrick Kelly. Oxford: Oxford University Press.
- Lapavistas, Costas. 2000. Money and the analysis of capitalism: the significance of commodity money. *Review of Radical Political Economics*, 32: 631-656.
- Locke, John. 1988 [1689]. *Two treatises of government*, ed. Peter Laslett. Cambridge: Cambridge University Press.
- Locke, John. 1991a [1691]. Some considerations of the consequences of the lowering of interest, and raising the value of money. In *Locke on money* [SG], vol. 1, ed. Patrick Kelly. Oxford: Oxford University Press.
- Locke, John. 1991b [1668]. Some of the consequences that are like to follow upon lessening of interest to 4 per cent. In *Locke on money*, vol. 1, ed. Patrick Kelly. Oxford: Oxford University Press.
- Locke, John. 1991c [1696]. Further considerations concerning raising the value of money. In *Locke on money*, vol. 2, ed. Patrick Kelly. Oxford: Oxford University Press.
- Mack, Eric. 2009. *John Locke*. London: Continuum.
- Macpherson, Crawford B. 1972. *The political theory of possessive individualism*. Oxford: Oxford University Press.
- Marx, Karl. 1977. Capital. In *Karl Marx: selected writings*, ed. David McLellan. Oxford: Oxford University Press.
- Nozick, Robert. 1974. *Anarchy, state and utopia*. Malden: Basic Books.
- Sreenivasan, Gopal. 1995. *The limits of Lockean rights in property*. New York: Oxford University Press.
- Steiner, Hillel. 1994. *An essay on rights*. Oxford: Blackwell Publishing.
- Tully, James. 1980. *A discourse on property: John Locke and his adversaries*. Cambridge: Cambridge University Press.
- Vallentyne, Peter, and Hillel Steiner. 2000. *Left-libertarianism and its critics: the contemporary debates*. Basingstoke: Palgrave Macmillan.
- Vaughn, Karen I. 1980. *John Locke: economist and social scientist*. Chicago: University of Chicago Press.
- Wade, Robert. 2006. The case for a global currency. *International Herald Tribune*, Editorials and commentary, 3 August.
- Wade, Robert. 2009. From global imbalances to global reorganization. *Cambridge Journal of Economics*, 33: 539-562.
- Waldron, Jeremy. 1988. *The right to private property*. Oxford: Oxford University Press.
- Weymark, John A. 1980. Money and Locke's theory of property. *History of Political Economy*, 12: 283-290.
- Wood, Neal. 1984. *John Locke and agrarian capitalism*. Berkeley and Los Angeles: University of California Press.

Luca Jacopo Uberti is a research fellow at the American University in Kosovo, where he teaches a course in political theory that focuses on contemporary theories of economic justice. His research interests straddle political economy and political philosophy, and he is currently benefitting from a research grant to study the political economy of trade and industrial policy in Kosovo.

Contact e-mail: <uberti@aukonline.org>

Hayek on the wisdom of prices: a reassessment

RICHARD BRONK

London School of Economics and Political Science

Abstract: This paper re-examines Hayek's insights into the problem of knowledge in markets, and argues that his analysis remains pertinent but has serious flaws. His central thesis—that the market price system is essential for communicating information and coordinating transactions wherever knowledge is dispersed and innovation renders the future uncertain—remains a potent explanation for the failures of central economic planning. His analysis that aggregate statistics necessarily abstract from contextual and tacit knowledge has important but widely ignored implications for the contemporary use of statistics in financial risk models. The recent financial crisis, however, shows that market prices can give very misleading signals for long periods, and it represents a key example of ways in which Hayek's thesis is incomplete. In particular, Hayek's analysis falls short by ignoring the role of dominant narratives, analytical monocultures, self-reinforcing emotions, feedback loops, information asymmetries and market power in distorting the wisdom of prices.

Keywords: F. A. Hayek, prices, knowledge, uncertainty, narratives, aggregate statistics

JEL Classification: A1, B31, B4, B5, D8

This paper re-examines Hayek's argument for what I call "the wisdom of prices". His thesis, that the market price system has a unique capacity to solve the problem of knowledge faced by economic agents, has always been provocative and contentious. Initially, this was because it challenged the very possibility of the central planning that was a central tenet of both socialist thought and policy practice in most western war-time economies. On this score, history has been kind to Hayek. But his thesis remains contentious today: first, because it throws doubt on the knowledge assumptions of the efficient markets and rational

expectations hypotheses, and suggests important problems with the use of statistics by credit rating agencies and the risk management departments of financial institutions; and, secondly, because faith in the “wisdom of prices” has been seriously challenged by the misleading signals given by prices in financial markets over the last decade or more.

In his famous paper “The use of knowledge in society”, Hayek attempted nothing less than to recast the central problem of economics as one of knowledge—of how society can make use of “knowledge which is not given to anyone in its totality” (Hayek 1948b [1945], 78). In a nutshell, he argued that the problem of economics had been miscast as how to achieve the efficient allocation of *given* factors on the basis of *given* data, with the corollary implication that correct foresight is at least theoretically feasible. This had in turn led to the mistaken belief that governments with superior aggregate statistics at their disposal could plan or intervene successfully to improve economic and social outcomes. In fact, Hayek argued, it is only through the unimpaired operation of the market and the signalling of the price system that we can discover the information about preferences, costs, requirements and market opportunities that we need to make good decisions. Such knowledge is otherwise often irremediably dispersed, subjective and tacit; or it may remain as yet undiscovered by anyone. Hayek invented one of the great metaphors of economics to explain the role of the price system in solving the problem of knowledge: he described “the price system as a kind of machinery for registering change, or a system of telecommunications which enables individual producers to watch merely the movement of a few pointers”; and he continued:

The marvel is that in a case like that of a scarcity of one raw material, without an order being issued, without more than perhaps a handful of people knowing the cause, tens of thousands of people whose identity could not be ascertained by months of investigation, are made to use the material or its products more sparingly; that is, they move in the right direction (1948b [1945], 87).

This argument for the wisdom of prices, and for the associated “*epistemological* impossibility” of socialist planning (Gray 1998, 40), is Hayek’s greatest achievement; and it is one that was thoroughly vindicated by the fate of the Soviet and Comecon systems, which manifestly failed to solve economic and social problems by planning in the absence of a market price system. As this paper will show, many of

Hayek's arguments for refocusing economics around the problem of knowledge and avoiding the misuse of aggregate statistics also remain highly pertinent to understanding contemporary failures in standard economic analysis and financial risk modelling.

Despite these crucial insights, however, the 2008 financial crisis highlights a need to re-assess Hayek's thinking and to consider the limits of its applicability. The crisis, and the years leading up to it, have revealed that the price system, even in a relatively free and liquid set of markets, can give profoundly distorted signals over a long period, and that market movements can themselves be deeply destabilising. Given the central importance of financial markets in the modern economy, and their strong association with free-market faith in the wisdom of prices, the financial crisis presents a serious challenge to Hayek's theory. We owe it to ourselves, and to Hayek's memory, to understand why his theory of knowledge and the epistemological role of prices has proved deficient in relation to modern financial markets despite its earlier prescient analysis of why planned economies cannot work. Some of the reasons highlighted in this paper for the failure of the price mechanism in the lead-up to the 2008 crisis are fairly specific to the operation of modern financial markets. But, I shall argue that other reasons discussed here represent more general qualifications to the applicability or validity of Hayek's theory. In particular, I shall argue that in conditions of uncertainty all markets are prone to being unduly influenced by homogenous group narratives that undermine the ability of market prices to reflect decentralised cognition in the way Hayek envisaged. My broader contention is that a series of lacunae in Hayek's thinking explain his failure to foresee how, in these and certain other circumstances, a belief in the wisdom of prices may prove misleading, or even self-defeating.

KNOWLEDGE AS THE CENTRAL PROBLEM OF ECONOMICS

When Hayek claimed that the central problem of economics is the "*division of knowledge*" (Hayek 1948a [1937], 50), he was consciously aping Adam Smith's analogous focus—the division of labour. Hayek's dialectical target, though, was the focus in contemporary mainstream economics on finding solutions to the problem of optimising among *given* preferences on the basis of *given* data. This, he argued, "is emphatically *not* the economic problem which society faces" (Hayek 1948b [1945], 77). The mainstream assumption that key data is

“given”—“at the command of everybody”—“disregards the fact that the method by which such knowledge can be made as widely available as possible is precisely the problem to which we have to find an answer” (Hayek 1948b [1945], 81). Hayek did not doubt the value of markets in achieving benign coordination, but for him the question was how this is achieved when economic agents are each operating with very little overall knowledge (Caldwell 2004, 336). Standard economics simply assumes away or ignores the central problem of knowledge. As Hayek put it:

The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess (1948b [1945], 77).

Hayek’s point here is more profound than the obvious truth that crucial information is dispersed across many individuals. Although difficult to solve in practice before the computer age, the challenge of the dispersion (or even quantity) of information presents few insuperable obstacles, in theory, to adequate and centralised knowledge. Instead, the main problems of knowledge in economics relate, in Hayek’s view, to the fact that the knowledge used by social agents is subjective (as well as dispersed) and often also tacit and subject to constant change.

Hayek gave a brilliant dissection of the confused term “given data”—a term still widely employed in economics. He pointed to an ambiguity between data in the sense of “objective real facts, as the observing economist is supposed to know them” and data, “in the subjective sense, as things known to the persons whose behaviour we try to explain” (Hayek 1948a [1937], 39). Hayek was in no doubt that the facts that matter in the social sciences are the opinions or “views held by the people whose actions we study” (Hayek 2010 [1952], 91). In other words, the data of the social sciences are “subjective” in the ontological sense that they “deal in the first instance with the phenomena of individual minds”—with opinions and perspectives that are necessarily incomplete and inconsistent; and it is these very opinions that constitute social reality by motivating action (Hayek 2010 [1952], 92, 99f).

Hayek was steeped in the Kantian tradition of seeing human knowledge as inevitably structured by mental concepts and categories that we furnish (Gamble 2006, 119). But we can, I think, discern three further steps in his argument that are more or less explicit in his use of the slippery term “subjective” in his discussion of knowledge. First, he inherited from Menger and others in the Austrian School a subjective theory of value, where “value is conferred on resources by the subjective preferences of agents and cannot be explained as an inherent property of any asset or resource” (Gray 1998, 16). This is important because only the individual concerned has full access to the value she places on goods or to her assessment of the opportunity cost of investments she makes. Secondly, since reality is multifaceted and complex, it cannot be grasped with any one central over-arching perspective: our views of the world are inescapably incomplete, perspectival and diverse. For Hayek, the term “subjective” is not a synonym for “erroneous”. Rather, he sees the illusion of objective and complete knowledge as what inevitably leads to error (Hayek 2010 [1952], 93). And thirdly, given his understanding of the subjective nature of value and the partial and perspectival nature of all knowledge, Hayek bought into the post-Kantian view that it is our particular interpretations and opinions that guide behaviour and (in part at least) construct social reality in their own image. Such is the stuff of economics.

The dispersed, subjective and perspectival aspects of knowledge are only part of the problem Hayek identifies. Equally important is that much of our knowledge is necessarily contextual and tacit—impossible to extract from “the particular circumstances of time and place” (Hayek 1948b [1945], 80). Tacit knowledge is the localized knowledge of *how* to do things, and it is knowledge that cannot easily be articulated and passed on to others in explicit and codified form, nor “processed by a committee or by a computer” (Hodgson 1999, 47, 60). Moreover, tacit knowledge implies, for Hayek, necessary limits to the reach of theory. As Gray (1998, 15) puts it: “theory is for him only the visible tip of the vast submerged fund of tacit knowledge, much of which is entirely beyond our powers of articulation”.

Finally, implicit in much of Hayek’s work is an acknowledgement of a yet more fundamental aspect of the problem of knowledge facing economic actors deciding how to act or invest for the future. Hayek viewed the market as a “discovery procedure”, but what actors need to discover is not limited to existing localized and tacit knowledge

available only to individuals. Buchanan and Vanberg (1994 [1991], 323, 328) point out that Hayek's language of discovery is somewhat unfortunate in that it might seem misleadingly to imply that the future—future alternatives, opportunities, costs, preferences and so on—is already “out there”, waiting to be discovered. Instead, they argue that the market is a “creative process”, characterized by constant innovation and novelty, and that the future does not exist ahead of its creation by the transformative power of this innovation and novelty. In other words, innovation implies a radical ontological limit to knowledge about the future. Shackle (1979, 52f) explained this perfectly, when he spoke of our “own original, ungoverned novelties of imagination [...] injecting, in some respect *ex nihilo*, the unforeknowable arrangement of elements”. The future, that is, cannot be known *ex ante* because it is still to be created by how “we imagine, will and choose it to be” (Bronk 2009, 219); and this first order uncertainty implied by innovation and our imagining of new possibilities is “compounded by uncertainty about the second-order creative reactions of others” (Bronk 2011, 9). In several passages, Hayek appears to grasp this most corrosive aspect of the problem of knowledge. For example, he speaks of much knowledge being “by no means ‘in existence’” in ready-made form, adding: “Most of it consists in a technique of thought which enables the individual engineer to find new solutions rapidly as he is confronted with new constellations of circumstances” (Hayek 1997a [1935], 95).

THE IMPOSSIBILITY OF CENTRAL PLANNING AND THE DANGERS OF AGGREGATE STATISTICS

These limits to knowledge form the core of Hayek's arguments against central planning. For him, the problem goes far beyond the manifest difficulties of amassing the required volume of existing dispersed information or making the requisite calculations to arrive at an optimal outcome, without the help of the market price mechanism. It follows from his analysis reviewed above that there are several interlocking reasons why successful central planning is impossible, and why even the advent of computer processing power could not save command economies (Hodgson 1999, 52-54). First, much of the required knowledge is tacit and cannot be made readily available in codified form to planners. Secondly, it is impossible for the knowledge generation capacities of all the divergent and incommensurable perspectives of the myriad market players in a complex and multi-faceted world to be

replicated by any *one* perspective or theoretical framework, however smart. Thirdly, it is impossible for the planner to know the subjective values that people, in all their variety, place on economic goods. But the killer fact, as Hayek stresses, is that we live in a world of constant flux. Our subjective “tastes change from moment to moment”; and our technical possibilities are constantly altering as we innovate and “discover” technical improvements in the face of new challenges. A centrally planned economy would not only have to allocate *existing* resources efficiently; it would also have to rival the knowledge generation and *discovery* capacities of the decentralised operation of the market mechanism. This would be difficult, since it is market competition itself that provides most of the incentives to adapt and innovate: “profits as an inducement to change cannot be dispensed with” (Hayek 1997a [1935], 95f, 108).

Hayek was singularly unimpressed with Lange’s attempt to counter these arguments by suggesting that the state could ape the price mechanism by acting the role of the Walrasian auctioneer and setting prices centrally by trial and error. Hayek argued that this proposal arose from an “excessive preoccupation with problems of the pure theory of stationary equilibrium”, and an under-appreciation of the need for interminable adjustments to new situations, new needs and new opinions. He continued:

With given and constant data such a state of equilibrium could indeed be approached by the method of trial and error. But this is far from being the situation in the real world, where constant change is the rule. Whether and how far anything approaching the desirable equilibrium is ever reached depends entirely on the speed with which the adjustments can be made (Hayek 1997b [1940], 123).

Perhaps the most topical aspect of Hayek’s criticism of the epistemological claims of socialists and central planners relates to their heavy use of aggregate statistics. There are two key elements to Hayek’s thinking on the misuse of statistics. First, he argued that the “blind transfer of the striving for quantitative measurements” from the natural sciences to the study of human relations—on the grounds that it is somehow more scientific than qualitative analysis—is “probably responsible for the worst aberrations and absurdities produced by scientism in the social sciences” (Hayek 2010 [1952], 114). Such an approach tends to ignore anything not easily measurable, abstracts from

differences between subjective assessments, and homogenises frames of reference—with a consequent inevitable loss of analytical and interpretive texture. It also abstracts from local contextual factors and any tacit knowledge that cannot be codified in data, “lumping together [...] items which differ as regards location, quality, and other particulars, in a way which may be very significant for the specific decision” (Hayek 1948b [1945], 83). As well as these perils of quantification, Hayek was also very wary of *aggregate* analysis in general and statistical *averages* in particular, arguing that they tend to obscure micro-level dynamics and give a misleading impression of greater stability in relationships over time than in fact exists.

This distrust of the knowledge content of aggregate statistics made Hayek almost as critical of macroeconomics as a discipline—and especially Keynesianism—as he was of central planning (Hayek 1967a [1962], 262). Hayek thought that attempts to use models relying on aggregate inputs to predict and manage demand in the economy tend to assume, as Gray (1998, 88f) puts it, “more in the way of concrete knowledge of the real relationships which govern the economy than any administrator could conceivably acquire”. It is fair to surmise that Hayek would have been equally critical of the modern risk management industry and credit rating agencies, had he lived to see them dominate financial markets and public policy. The way in which rating agencies aggregate information on corporate and national entities operating in complex dynamic situations to provide a centralised assessment of risk that can replace decentralised market cognition can be seen as analogous to the efforts of central planning bureaux that Hayek so despised. Similarly, the fact that large banks seek to codify, quantify and aggregate the variables they face in an uncertain environment and in a myriad of different contexts, and reduce these to summary Value at Risk metrics, runs counter to Hayek’s strictures on the dangers of abstracting from the localised, tacit and constantly changing knowledge of individual agents. The financial crisis has shown that both these attempts at aggregating information tend to give an illusion of control, while failing to reflect key factors in dynamic situations.

THE WISDOM OF PRICES AND THE MARVEL OF MARKET COORDINATION

If central agency statisticians cannot solve the problem of knowledge, how does the free market either solve it or, alternatively, sidestep the need to do so? As Hayek points out, it cannot solve it simply by

devolving decisions to individuals with access to their own subjective, local and tacit knowledge. There also needs to be a mechanism that allows the person on the spot to acquire enough information about the requirements, subjective beliefs and expectations of others to be able to coordinate her actions with everyone else's. This is where the wisdom of prices comes in: it is prices that "act to co-ordinate the separate actions of different people"; and they do so because the price system acts as "a mechanism for communicating information", a role it performs with great epistemological economy (Hayek 1948b [1945], 84-86). As Hayek wrote:

The most significant fact about this system is the economy of knowledge with which it operates, or how little the individual participants need to know in order to be able to take the right action. In abbreviated form, by a kind of symbol, only the most essential information is passed on and passed on only to those concerned (1948b [1945], 86).

In this very special sense then, a kind of knowledge—which can be shared in summary form by anyone who needs it—is an emergent phenomenon from the continuous process of market interaction. Thanks to the information conveyed by prices, individual agents can act with the benefit of a type of wisdom that is digestible and yet more comprehensive than they alone could otherwise acquire or even understand. This wisdom of prices is a product of the myriad of coordinated pricing decisions by individuals, where each decision is made by individuals combining their own contextual and subjective knowledge with the outline messages they glean about the views of others as expressed by the prices they in turn are willing to accept or pay. In this way, as Vernon Smith (2008, 105) puts it, prices are "both the carriers of information and the result of that message exchange".

We will come back to some reservations about the knowledge content of market prices later in the paper, but at this stage it is worth exploring further how Hayek envisages the price system operating. First, he argues that prices do more than convey key information about the beliefs of others; they also direct our attention "to what is worth finding out about market offers" (1978a [1968], 182). That is, price movements may grab our attention and alert us to areas that others find of interest or concern, prompting further research of our own (Shiller 2005, 171). Secondly, it is from prices that we discover the existence of innovative

new developments. By way of example, Hayek argues that cost curves are not “objectively given facts”. Rather, the cheapest method is something “which has to be discovered, and to be discovered anew sometimes almost from day to day”; and it is the price system that acts to communicate the discovery because new ideas announce themselves by innovators competitively undercutting the prices of established producers (Hayek 1997b [1940], 130). It is exactly this sort of informational role that prices are only able to perform when the market is not impaired by government intervention. In *The road to serfdom*, Hayek argues:

Any attempt to control prices or quantities of particular commodities deprives competition of its power of bringing about an effective co-ordination of individual efforts, because price changes then cease to register all the relevant changes in circumstances and no longer provide a reliable guide for the individual's actions (1944, 27).

This analysis has largely been vindicated by the negative experiences of those economies that have grossly interfered with the price system by using price controls or rationing. But, as we shall see, Hayek mistakenly ignored the possibility that there might be endogenous market influences that could similarly undermine the wisdom of prices without government interference playing a deciding role.

HAYEK'S CHALLENGE TO STANDARD ECONOMICS AND THE ROLE OF GOVERNMENT

Despite Hayek's argument for the wisdom of prices and his paeon of praise for free markets, several aspects of his writings make him a distinctly uncomfortable figure for mainstream economists. In the first place, Hayek became increasingly critical of the emphasis in economic models on a static conception of efficiency, and indeed of the very notion of an optimal equilibrium (Hayek 1997b [1940], 123; Gamble 1996, 69). As Hayek wrote:

Economists usually ascribe the order which competition produces as an equilibrium—a somewhat unfortunate term, because such an equilibrium presupposes that the facts have already all been discovered and competition therefore has ceased (1978a[1968], 184).

Hayek may have argued that prices reflect key information and register relevant changes in circumstances, but there is much in his

writings to suggest that he would have been dismissive of modern variants of the efficient markets hypothesis. As well as downplaying the notion of optimal equilibrium, he never argued that prices themselves could reflect *all* relevant information. For Hayek the role of prices was to *supplement* each agent's particular cognition (and local knowledge) with a summary reflection of all the market decisions that others have made in the light of their respective particular perspectives and circumstances. Hayek would have been equally critical of the rational expectations hypothesis and its central assumption that, thanks to the competitive elimination of systematic errors, the representative agent internalises the correct theory of the economy (Frydman and Goldberg 2011, 67, 91). There can be no representative agent in the Hayekian world of radically decentralised knowledge, diverse perspectives and subjective valuations; there can be no single theory that encapsulates tacit and contextual knowledge and all the relevant aspects of our complex world; and there is no single future optimal equilibrium 'out there' on which all rational expectations must converge. Instead, the future we face has yet to be created by discoveries we may *ex ante* not even know we need.

In many ways, Hayek's work prefigures modern complexity theory in its epistemological challenge to standard economics. While he lacked the sophisticated non-linear mathematics and agent based modelling employed by many complexity theorists today; and while he sometimes appears to have confused complex in the sense of 'complicated' (i.e., a large number of variables and aspects) with the more technical sense of the dynamic emergence of novel outcomes; there is something very radical in Hayek's conception of the market as a "spontaneous order" or "catallaxy" that emerges in an unplanned way from the interaction of heterogeneous agents (each endowed with only partial knowledge) following abstract rules that determine only general patterns of behaviour (Hayek 1967b [1964], 27; Hayek 1967c [1965], 92; Hayek 1978a [1968], 183). The most challenging element of Hayek's theory of markets as "complex phenomena" is his insistence that their complexity renders precise prediction impossible. In contrast to Friedman's insistence that economics should be a positive science, judged by the "precision" and "scope" of its falsifiable predictions (Friedman 1994 [1953], 181), Hayek argued that economists should be content with a lower degree of explanation and a lower 'degree of falsifiability'. And, rather than attempt precise predictions, they should engage merely in

“explanations of the principle” and “pattern prediction”. In other words, they should use a scientific method more like that employed by evolutionary biologists than by astro-physicists (Hayek 1967b [1964], 22-31).

Once economists renounce the assumption that agents can optimise on the basis of *given* factors and correct foresight (an assumption which allows precise prediction); and once they acknowledge that we operate in complex and dynamic systems where we can never know “all the circumstances which will determine the outcome”; they are forced to accept limits to prediction (Hayek 1978b [1974], 24, 27). And with limits to prediction come limits to our ability to control outcomes by using theory-based knowledge. As Hayek said in his Nobel address, “to entrust to science—or to deliberate control according to scientific principles—more than scientific models can achieve may have deplorable effects” (1978b [1974], 30). It is likely that many of today’s market participants, faced with the serial failure of economists to predict outcomes with any precision, would acknowledge Hayek’s insights on the limits to prediction. It is less clear how far they or the body politic have accepted the corollary limits to control, and acknowledged the dangers, for example, of expecting central banks and governments to engineer a smooth glide path to recovery on the basis of economic models. There is little doubt that, if alive, Hayek would have laid some of the blame for the recent financial crisis on repeated interventions by the Federal Reserve to limit asset price corrections from 1998 onwards—the so-called “Greenspan put”—which had the unintended consequence of fuelling an unsustainable boom in credit; and little doubt, too, that he would have been queasy about current quantitative easing policies.

Hayek acknowledged that market prices alone do not provide market agents with all the information they need; and, indeed, he recognised a greater role than most modern economists do (outside the Institutionalist school) for rules of conduct and institutions as carriers of both tacit knowledge and the wisdom generated from the trials and errors of the past. Such rules are “the product of a slow process of evolution in the course of which much more experience and knowledge has been precipitated in them than any one person can fully know” (Hayek 1967c [1965], 92). For Hayek, the great error of the modern age was to assume that we could do without these evolved rules or traditions, and instead use rational (economic) models to engineer a better future.

MISSING ELEMENTS: THE ROLE OF NARRATIVES AND EMOTIONS

At this point in the lecture on which this paper is based, I sensed some in my audience becoming restive. Surely, they might reasonably have objected, it is precisely a Hayekian belief in the wisdom of market prices that is responsible for the economic and financial mess in which we have found ourselves since 2008. Is it not belief in the epistemic and other virtues of an unimpeded market mechanism that led to the thirty-year experiment with deregulation which has, paradoxically, swept away the very traditional rules and institutional repositories of wisdom that Hayek valued (Gray 1998, 153-155)? And, when it comes to the credit crunch, is it not clear that prices gave grossly misleading signals of value for a considerable period and failed to alert relevant actors to the problems brewing until too late? It is time, therefore, to examine some of the main factors that caused financial market prices to be so distorted, and to assess how far these suggest general qualifications to Hayek's theory and link to broader problems in his conception of the wisdom of prices.

For all his "epistemological pessimism" (Gamble 2006, 118), Hayek underestimated the degree of radical uncertainty we face and oversimplified the way we cope with it. As Keynes (1936, 149) argued: "The outstanding fact is the extreme precariousness of the basis of knowledge on which our estimates of prospective yield have to be made". But even Keynes did not theorise fully about the various causes of uncertainty and its crucial link to innovation and new ideas (Bronk 2009, 215; Dunn 2003, 177). This was left to Shackle, who noted: "What does not yet exist cannot now be known. [...] [We] cannot claim Knowledge, so long as we acknowledge Novelty" (1992 [1972], 3, 26). It is precisely because the market is a machine for generating innovation and novel ideas that we face deep "ontological uncertainty" (Bronk 2011, 8f).¹ Moreover, this is not a problem of dispersed information but one of

¹ Under the umbrella of Keynesian uncertainty, an important distinction can be made between "epistemological uncertainty", where relevant probabilities are in practice unknown (because of the difficulty of computing all relevant parameters), and "ontological uncertainty", where probabilities are *ex ante* logically unknowable (because of an indeterminacy at the level of reality implied by innovation and novelty) (Bronk 2011, 8-11; Skidelsky 2009, 88). While limited progress can be made in reducing epistemological uncertainty, ontological uncertainty presents intractable barriers to knowledge: it may be simply impossible to know even the categories or entities that will comprise future reality since many of them have yet to be invented (Bronk 2013).

“symmetric ignorance” (Skidelsky 2009, 45): the inventor of a novel product may know a little more than others about its potential, but even he is largely unaware of the implications of its adoption by other innovative and resourceful agents.

When faced with ontological uncertainty, everyone is feeling his way forward with little firm indication of what the future will bring. With no single correct version of the future out there to anchor expectations, and a vast space of possible outcomes, prices reflect not so much our decentralised *knowledge* as the way we *imagine* or *hope* the future will be: “Valuation is expectation and expectation is imagination”, as Shackle (1992 [1972], 8) epigrammatically expressed it. Crucially, these imaginings generally take the form of narratives, which provide us with scripts that “keep ontological uncertainty at bay” (Lane and Maxfield 2005, 4). It is the stories we tell ourselves that help us to chart our way through the unknown future and interpret the constant flow of new information, and that provide us with “rationales to support action” (Tuckett 2011, 160). It is stories that “provide parameters for decision-making [...] *despite* the uncertainty” we face (Beckert 2011, 5), and thereby help us decide the prices we are willing to pay. In consequence, as Tuckett (2011, 24) shows in his empirical findings, the valuation of assets “is inextricably linked with the stories people tell about their futures”, and the emotions that attach to these narratives.

This vital role for imagination and narratives in conditions of uncertainty does not in itself invalidate Hayek’s faith in the (relative) wisdom of prices. So long as prices are a function of *heterogeneous* perspectives and the multiple imaginings, dreams and narratives of all market participants, they may still help us spot emerging patterns better than we otherwise could (Bronk 2013). But, as Keynes reminded us, in conditions of uncertainty investors normally resort to “conventional” methods of valuation, and are affected by “mass psychology” and “waves of optimistic and pessimistic sentiment” (Keynes 1936, 152-154). In other words, in conditions of uncertainty, prices tend not to reflect the *decentralised* cognition that Hayek believed to be the main epistemic advantage of the price system, but rather *group* emotions and *shared* conventions or narratives.

It is a feature of human beings that peer pressure affects the way we assess evidence (Cassidy 2009, 188f); and that, because we are social animals, we tend to think similarly and to be influenced—however independent we believe our outlooks to be—by “a *Zeitgeist*, a spirit of

the times” (Shiller 2005, 157). Each period of history tends to be marked by what Foucault called “totalising discourses” or Lyotard “grand narratives” (Drolet 2004, 20, 25). And, when faced with the uncertainty caused by innovation, we are prone to cope by adopting shared “new era stories” that inspire confidence and replace the “stories of the past” (Akerlof and Shiller 2009, 55). The resulting homogeneity of narratives and beliefs, and related tendency to “*groupfeel*”, undermines the cognitive diversity essential to healthy markets (Tuckett 2011, 19). Because any single narrative, model or perspective shines a light only on certain aspects of what is going on, reliance on one narrative may induce blindness to factors which are unforeseen or ignored by that narrative. And whenever we *all* rely on *one* perspective—on a *single* dominant narrative, modelling framework or gut instinct—we all tend to focus on the *same* aspects of what is going on and to share the *same* cognitive myopia (Bronk 2011, 14f). When that happens, we can expect prices to reflect our collective bias and to become detached from any fundamentals that our shared perspective ignores.

This argument for the cognitive and price distortion effects of homogenous new-era stories and group emotions is potentially germane to any market where innovation causes widespread uncertainty—the information technology sector being a good example. This article, though, focuses on the argument’s obvious relevance to explaining why market prices proved misleading before the 2008 financial crisis.² The decade leading up to the crisis was characterised by massive financial and policy innovation, which led to high levels of ontological uncertainty (Bronk 2011, 11). In such conditions, investors duly relied on convention and new era stories, which in turn engendered widespread confidence. As Power (2007, viii, 74f) argues, financial markets became structured by a “world-level grand narrative of risk management”, which fostered an illusion of control by confusing radical ontological uncertainty with measurable risk; and nearly all players saw it as best practice to use Value at Risk models to assess the risks they were running. Indeed, as Haldane (2009, 4) notes, the pervasive rhetoric was that we had entered a new era of “simultaneously higher return and lower risk” as a result of

² I am assuming it is accepted that market prices were misleading before the 2008 market crash: they could only have been a good reflection of fundamentals at both pre- and post- crash market levels on an extreme view that the huge price changes reflected only random exogenous shocks that could not *ex ante* have reasonably been anticipated or spotted by anyone.

“a shift in the technological frontier of risk management”. Such an analytical monoculture, combined with an overlapping narrative of efficient markets (where prices reflect all available information), caused most players to be over-confident and to miss or ignore warning signs that in retrospect were obvious. Worse still, the risk modelling monoculture (and certain exogenously determined global regulatory frameworks) led most players to share similar beliefs; and since beliefs structure action, this caused very high correlations in behaviour (Bronk 2011, 15). Prices responded accordingly and the wisdom of prices was hopelessly compromised.

Hayek would not, I think, have expected this to happen. In part, that is because his “intransigent methodological individualism” (Skidelsky 2006, 95), and his attack on “methodological collectivism”, made him unduly wary of explanations at the level of group dynamics or “social phenomena” (Hayek 2010 [1952], 117f). Perhaps for this reason, his subjective account of knowledge largely ignores the all-important *inter-subjective* or *social* construction of motivating beliefs. It is true that Hayek was well aware that individual cognition is structured by language, institutions and abstract rules. But in his analysis these are not short or medium-term contingent social factors but the product of a long-term process of social evolution that winnows out misleading rules and ensures they are superior aids to cognition rather than potentially misleading frames (Gamble 1996, 54; Gray 1998, 41, 141).

There are two weaknesses in Hayek’s evolutionary conception of institutions: first, it assumes that the fitness landscape of tomorrow will be similar to that of yesterday despite radical innovations in the way we operate; and secondly, it ignores the social power of apparent confirmation of an ultimately flawed narrative or rule by any medium-term success it has in creating reality in its own image. In other words, we should not assume that group narratives that will ultimately prove misleading are selected out by competitive markets, since future challenges to the validity of a narrative or theory may be different from those of the past or present; and, over the medium-term horizon on which we operate, there is a strong tendency for belief in a narrative (that may in the long-run be flawed) to be reinforced by the impact on market prices of many players adopting that narrative. Let us consider such endogenous mechanisms for self-reinforcing error further.

FINANCIAL MARKET MECHANISMS FOR SELF-REINFORCING COGNITIVE HOMOGENEITY AND ERROR

In conditions of uncertainty, markets become “markets in stories”—markets in competing interpretative narratives (Tuckett 2011, 159). There are always several plausible narratives about the future yet to be created by innovative agents, and the outcome of this competition in interpretive narratives is indeterminate and partly a function of intentional strategy and rhetoric. Competing narratives are also needed to help make sense of the movement of prices, since prices themselves usually require interpretation before they can be used to make decisions. This ambiguity in the meaning of prices is partly due to their being at best only an economical symbol of existing tacit and decentralised knowledge and judgements;³ and partly because prices reflect the shifting group narratives with which we interpret our uncertain future predicament. Traders often ask, for example, what a move in the oil price means, and look for a narrative that makes sense of unusual movements. When they have found one they like, and acted accordingly, a self-reinforcing dynamic may take hold if the traders then try to convince others to adopt their preferred narrative, in order to validate their market decisions and investments. Policy-makers are another intentional source of shared narratives, as they seek to guide our expectations. Holmes (2009, 385f) argues, for example, that a key part of a central bank’s armoury is the use of persuasive and “skilfully composed narratives” that “serve as an analytical bridge to the near future” and align expectations with an inflation or growth target. This strategic use of narratives is often very effective because, at least to some extent, the narratives and models we use are “performative”—that is, they succeed in shaping reality in their own image by structuring the beliefs that motivate action. As Beckert (2011, 8) puts it: “If a sufficient number of investors believe in the fictional depiction it becomes a self-fulfilling prophecy”.

This performative attribute of narratives or models may, however, create an additional problem of knowledge, for market participants and

³ Hayek saw the very economy of the information provided by prices as a virtue; but it is also a weakness. Prices abstract from nuances of interpretation, giving only a headline reading of the supply and demand that is generated in the light of decentralized knowledge and social narratives. As Westbrook (2004, 51) puts it provocatively: “Money is structurally incapable of transferring much information, as a language composed of a single word would be”. Holmes (2009, 410) argues that, consequently, we need to supply narratives that identify “what the act of pricing discards or suppresses”.

economist alike. To the extent that (for social, strategic or rhetorical reasons) a narrative becomes dominant and governs the beliefs and therefore the actions of most agents, prices respond accordingly; and this price response may (if sufficiently in line with the narrative) reinforce confidence in the veracity of the narrative, with a further knock-on impact on prices—an example of what Soros (2010, 14) calls “reflexive feedback loops”. But such reflexive lock-in happens despite the fact that most narratives or models have weaknesses and miss some aspects of what is taking place. The difficulty is that we cannot know *ex ante* whether the apparent short or medium-term confirmation of a narrative by price movements reflects genuine fulfilment of the narrative’s script or instead masks underlying problems with the narrative that have yet to become apparent. This new problem of knowledge caused by the performativity of dominant narratives would be easier to solve if most market practitioners retained access to the “generative friction” of using “multiple evaluative principles” and so remained receptive to anomalies that challenged their preferred narrative or interpretation of events (Stark 2009, 16f). But it is exactly this cognitive diversity and receptiveness that is compromised when a market as a whole is governed by an analytical monoculture—whether this monoculture is caused by a general conviction that there is one best practice (Bronk 2011, 17) or by a particular market narrative becoming dominant thanks to reflexive reinforcement by corroborating medium-term price movements.

Several social and technical features of markets can help reinforce such homogeneity of belief and increase the dangers of self-reinforcing error. First, as Akerlof and Shiller (2009, 56) argue, stories and the emotions of confidence and pessimism that attach to them spread “like viruses”; and, indeed, their transmission can best be modelled with the techniques of epidemiology. Market panic (or confidence) and associated narratives are self-reinforcing; and their spread (like that of epidemics) is unpredictable and given to sudden threshold effects. Secondly, Beunza and Stark (2012, 383-413) show that some traders have developed “reflexive modelling” that infers the views of other traders from the prices they pay in order to use these inferred views “as inputs to their own decision-making”. Beunza and Stark argue that this technique may aid “distributed cognition” and the “interplay of internal and external estimates” so long as the opinions and perspectives so inferred are heterogeneous, but that the same technique may lead to a

dangerous form of “cognitive interdependence” and serve to reconfirm error if most players initially share the same faulty analytical frame. A third factor—in this case exogenously determined by regulators—is the increasingly widespread international use of “mark to market accounting”, which requires that changes in prices be immediately reflected in balance sheet values. This leads to the danger of a “loss-spiral” if banks are forced to off-load other assets to make up for any loss before they would otherwise have needed to (Cassidy 2009, 309f). Such self-reinforcing price movements can be very destabilising. Even if the initial price signal is valid, there is a danger that, as a result of a shared accounting convention, correlations in the reactions to this price movement are unnecessarily high, and prices move far beyond the level that fundamentals would otherwise imply.

Another dynamic in financial markets that can lead to self-reinforcing errors was first articulated by Keynes (1936, 156) in his analogy of financial markets to a beauty contest where the prize goes to whoever correctly anticipates “what average opinion expects the average opinion to be”. In part, this dynamic is caused by the fact that, when the future is uncertain, there is an incentive for an investor to second-guess the opinions of others, not for their imputed informational value, but because opinions are reflected in prices with the result that short or medium-term market movements can be predicted (and a fortune made) by correctly anticipating the trend of average opinion. It is also, as Keynes (1936, 157f) wrote, because it is usually safer in career terms to be successful in such short-term momentum trades, or else to “fail conventionally”, than to risk an unconventional approach. For, an unconventional approach often implies short-term losses and may not be profitable for the investor before doubting clients or employers have terminated her career.

Orléan (2012, 316f, 331) takes Keynes’s argument a stage further, arguing that a financial market acts as a cognitive machine for producing “a reference opinion”—“an expression of ‘what the market thinks’”; and that this market opinion is a by-product of the self-referential process identified by Keynes, which renders agents “extremely attentive to the way in which the collective opinion is formed”, that is, to salient models and dominant narratives. Here then is a self-reinforcing market mechanism that drives the serial emergence of widely held market opinions or market-wide conventions. When this happens, I would argue, Hayek’s reasons for favouring the price system

as a generator of knowledge are threatened: market practitioners may no longer rely primarily on their diverse independent perspectives and interpretations of events or prices but may instead continually second-guess, and then rely exclusively on, emerging average opinion about what prices signal and the future holds—average opinion which may turn out to be misguided in the long run.

What is particularly paradoxical for Hayek's theory is that market practitioners often behave in this way, not merely to anticipate market moves speculatively, but because they have internalised a simplified version of Hayek's own message about the wisdom of prices. Largely under the auxiliary influence of the (non-Hayekian) efficient markets hypothesis that prices reflect *all* available information, a generalised belief in the wisdom of prices has become divorced from Hayek's insistence on each agent complementing the message from prices with his own (decentralised) cognition and local knowledge. In other words, while Hayek's central argument was that the price system is *necessary* for wisdom, market participants have come to believe the stronger claim that the information gleaned from prices is *sufficient* for wisdom—a view Hayek never shared. When they adopt this stronger claim, market actors believe it is rational to economise on their own analysis and follow the judgement of the herd as expressed in market prices, on the assumption that market prices give correct signals; and such an unqualified belief in market opinion may eventually cause persistent mispricing.

This perverse dynamic is similar to the phenomenon known as an “information cascade”, where people with limited knowledge follow the decisions of others rather than make independent judgements, on the frequently false assumption that the herd is likely to be right, and consequently end up converging on bad outcomes (Cassidy 2009, 189-191; Surowiecki 2004, 53f). A generalised belief in the wisdom of prices may be self-defeating if it encourages market participants to substitute emergent market opinion for their own independent judgments and to free ride on the wisdom of others as reflected in prices. When that happens, the price system no longer reflects the cognitive diversity and localised knowledge of all market agents, and instead it becomes an unstable product of reflexive interaction and social learning. Nor is this only a phenomenon in financial markets. Widespread belief that the wisdom of prices is sufficient (like widespread belief in the “wisdom of

crowds”⁴) may encourage in any market a tendency that at group level is self-defeating—a tendency for individuals to assume they can economise on their own localised knowledge and cognitive effort, thereby depriving the price system of much of its informational input.

MARKET POWER, INFORMATION ASYMMETRY, EXTERNALITIES AND THE QUALITY OF PRICES

Most of the mechanisms for entrenching cognitive homogeneity and self-reinforcing error considered above are particularly damaging to the wisdom of prices in the conditions of ontological uncertainty that are especially (though not exclusively) associated with rampant financial market innovation. There are, though, other factors that frequently impinge on the validity of Hayek’s argument for the epistemic value of the price system even in more stable environments and in non-financial markets.

The first is the distortion implied by gross inequalities of wealth and market power. Hayek’s assumption that key implications of decentralised knowledge and the full variety of subjective assessments are reflected in market prices (and that this is the main reason for the superiority of markets over central planning) ignores the problems implied by such inequalities. The degree of influence that any player has on market prices is a function of their market power and wealth; and this means that if those who possess some key decentralised knowledge have very little wealth or market power, and are outbid by those ignorant of the facts, the true picture may not be well reflected in prices. It is usually large players that, regardless of the superiority (or otherwise) of their knowledge, control the market prices on which others base their strategy. Market success often goes to the wealthy rather than the wise. Of course, inequalities in wealth and market power compromise more than the epistemic role of the price mechanism: they also compromise its moral claim to reflect the various subjective preferences and valuations of all market participants. Hayek argued that an advantage of the market mechanism is that, unlike central planning,

⁴ Surowiecki’s argument (2004) for the “wisdom of crowds” resembles Hayek’s for the wisdom of prices. Surowiecki argues that crowd decisions are wise when they aggregate a diverse set of decentralised and independent judgements. But Surowiecki (2004, 43f) also points out that generalised belief in the wisdom of crowds can be self-defeating: “[If] the group usually knows best (as I’ve argued it often does), then following the group is a sensible strategy. The catch is that if too many people adopt that strategy, it stops being sensible and the group stops being smart”.

it does not impose a single scale of values but instead reflects the full variety of preferences found in a free society (Hayek 1997b [1940], 138). But if wealth or market power is too concentrated then market prices largely reflect the subjective preferences of the rich.

Another equally serious lacuna in Hayek's thinking is his minimal recognition of the damage done to the wisdom of prices by market failures. These failures include the problem of *externalities*—the fact, for example, that the full *social* costs of pollution, congestion or resource depletion are rarely reflected in the prices that individual firms or consumers face in the market. They also include the problem of *information asymmetries* where one party to an exchange has an information advantage over the other. In such circumstances, as is well articulated in modern information economics, there may be opportunistic behaviour by the advantaged party. The fear of this alone is enough, in the absence of trust, to cause “thin” markets and the mispricing of deals (Akerlof 1970); and there is little doubt this problem played a part in the recent distortion of price signals in financial markets (Cassidy 2009, 164f). Crucially, though, the very tacit and contextual knowledge that Hayek believed to be an important reason for favouring the decentralised cognition of market pricing implies information asymmetries; and these asymmetries in turn imply that tacit knowledge is likely to be associated with thin markets characterised by distrust, mispricing and low liquidity. Hayek never, I believe, recognised this problem with his theory. In practice, any sector where tacit knowledge is key to the valuation of products (for example, the specialised mechanical engineering sector in Germany) needs non-market coordination mechanisms that allow relevant parties to share tacit knowledge and build trust (Bronk 2009, 162f).

A related problem with Hayek's theory concerns the pricing of non-standard products. Non-standard products are for Hayek a prime example of where tacit and contextual knowledge is key. But modern theory suggests that in these cases we do not have the level of liquidity and repeat transactions that allow prices to gain acceptance as fair public indicators of information. MacKenzie (2012, 336f, 345), for example, notes that in financial markets prices typically provide good information that most market participants are willing to trust only where there are standardised products, highly liquid markets, and continuous trading. Mackenzie argues that, additionally, there must be few concerns (arising from opportunism) about the “quality” of prices—

their “fairness”. This implies an important general qualification to Hayek’s theory: in the case of non-standard product areas and especially one-off transactions (where tacit and contextual knowledge and information asymmetries are often crucial) there is unlikely to be widespread faith in the fairness of posted prices. In these cases, prices are likely to be agreed upon only by parties who share crucial information through non-market mechanisms, and then the price mechanism itself is not a primary source of the knowledge required to transact. Furthermore, when prices for non-standard transactions are agreed in this way, they may never be made publicly available in a form that third parties can learn from.

CONCLUSION

As a critique of attempts to aggregate information centrally with statistics, Hayek’s analysis of the problems of knowledge remains peerless. His championing of the feats of coordination enabled by the market price mechanism also remains convincing in explaining why in most markets we rarely experience widespread gluts or shortages of products, and why we largely succeed in catering for an astonishing variety of subjective preferences. Such coordination makes use of knowledge that is never available centrally and it does so through the signals given by prices.

It is important, however, to recognise that, particularly in the conditions of uncertainty caused by widespread innovation (where no strong anchor for expectations exists), prices may be seriously misleading. In these conditions, market participants tend to gravitate to group narratives to help make sense of their predicament, with the result that prices tend to reflect a narrow range of partial perspectives rather than the fully decentralised and diversified cognition that Hayek correctly saw as key to the wisdom of prices. There are several mechanisms endogenous to financial markets that particularly favour the epistemologically dangerous emergence of *homogenous* frames of reference, widespread *conventional* opinions and analytical *monocultures*—notably feedback loops between economic (or modelling) narratives and prices. Paradoxically, another threat to the wisdom of prices is that widespread belief in the wisdom of prices can become self-defeating if it causes market participants (contrary to Hayek’s advice) to economise on their own decentralised cognition and free-ride on the wisdom of others, thereby depriving the price system of some

of its informational input. And, finally, Hayek's followers need to acknowledge the serious damage done to the knowledge-generating properties of market prices whenever markets are characterised by large inequalities of wealth, externalities or information asymmetries.

Like most great theories in the history of ideas, Hayek's theory of the wisdom of prices may be as interesting for the reasons we now want to qualify its applicability as for its continued insights in other areas. For, when we understand better the sources of weakness in the epistemic role of the price system, we may have a better chance of avoiding the conditions that generate widespread distortion of market prices. In particular, the argument advanced in this paper suggests the importance of safeguarding the cognitive diversity of market participants by minimising as far as possible endogenous market pressures for analytical monocultures and homogeneity of belief. It also suggests that great care is needed in weighing up the advantages of regulatory harmonisation against the disadvantages for the operation of the price mechanism of this exogenous source of cognitive homogeneity. But consideration of this regulatory trade-off is another story for another paper.

REFERENCES

- Akerlof, George A. 1970. The market for "lemons": quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84 (3): 488-500.
- Akerlof, George A., and Robert J. Shiller. 2009. *Animal spirits: how human psychology drives the economy and why it matters*. Princeton: Princeton University Press.
- Beckert, Jens. 2011. Imagined futures: fictionality in economic action. *MPIfG discussion paper*, No. 11/8. Max Planck Institute for the Study of Societies.
- Beunza, Daniel, and David Stark. 2012. From dissonance to resonance: cognitive interdependence in quantitative finance. *Economy and Society*, 41 (3): 383-417.
- Bronk, Richard. 2009. *The romantic economist: imagination in economics*. Cambridge: Cambridge University Press.
- Bronk, Richard. 2011. Uncertainty, modelling monocultures and the financial crisis. *The Business Economist*, 42 (2): 5-18.
- Bronk, Richard. 2013. Epistemological difficulties with neo-classical economics. *Review of Austrian Economics: Special issue: the many faces of neoliberalism* (forthcoming).
- Buchanan, James. M., and Viktor. J. Vanberg. 1994 [1991]. The market as a creative process. In *The philosophy of economics: an anthology*, ed. Daniel M. Hausman. Cambridge: Cambridge University Press, 315-335.
- Caldwell, Bruce. 2004. *Hayek's challenge: an intellectual biography of F. A. Hayek*. Chicago: Chicago University Press.
- Cassidy, John. 2009. *How markets fail*. London: Allen Lane.
- Drolet, Michael. 2004. *The postmodernism reader: foundational texts*. London: Routledge.

- Dunn, Stephen P. 2003. Keynes and transformation. In *The philosophy of Keynes's economics: probability, uncertainty and convention*, eds. Jochen Runde, and Sohei Mizuhara. London: Routledge, 170-181.
- Friedman, Milton. 1994 [1953]. The methodology of positive economics. In *The philosophy of economics: an anthology*, ed. Daniel M. Hausman. Cambridge: Cambridge University Press, 180-213.
- Frydman, Roman, and Michael D. Goldberg. 2011. *Beyond mechanical markets*. Princeton: Princeton University Press.
- Gamble, Andrew. 1996. *Hayek: the iron cage of liberty*. Cambridge: Polity Press.
- Gamble, Andrew. 2006. Hayek on knowledge, economics and society. In *The Cambridge companion to Hayek*, ed. Edward Feser. Cambridge: Cambridge University Press, 111-131.
- Gray, John. 1998. *Hayek on liberty*. 3rd ed. London: Routledge.
- Haldane, Andrew G. 2009. *Why banks failed the stress test*. Speech given at the Marcus-Evans Conference on Stress-Testing 9-10 February 2009. Bank of England. Website: <http://www.bankofengland.co.uk/publications/Documents/speeches/2009/speech374.pdf> (accessed April 23, 2013).
- Hayek, F. A. 1944. *The road to serfdom*. London: Routledge.
- Hayek, F. A. 1948a [1937]. Economics and knowledge. In *Individualism and the economic order*. Chicago: Chicago University Press, 33-56.
- Hayek, F. A. 1948b [1945]. The use of knowledge in society. In *Individualism and the economic order*. Chicago: Chicago University Press, 77-91.
- Hayek, F. A. 1967a [1962]. The economy, science, and politics. In *Studies in philosophy, politics and economics*. London: Routledge, 251-269.
- Hayek, F. A. 1967b [1964]. The theory of complex phenomena. In *Studies in philosophy, politics and economics*. London: Routledge, 22-42.
- Hayek, F. A. 1967c [1965]. Kinds of rationalism. In *Studies in philosophy, politics and economics*. London: Routledge, 82-95.
- Hayek, F. A. 1978a [1968]. Competition as a discovery procedure. In *New studies in philosophy, politics and economics and the history of ideas*. London: Routledge, 179-190.
- Hayek, F. A. 1978b [1974]. The pretence of knowledge. Nobel Memorial Prize Lecture. In *New studies in philosophy, politics and economics and the history of ideas*. London: Routledge, 23-34.
- Hayek, F. A. 1997a [1935]. The present state of the debate. In *The collected works of F. A. Hayek: socialism and war*, ed. Bruce Caldwell. Indianapolis: Liberty Fund, 89-116.
- Hayek, F. A. 1997b [1940]. Socialist calculation: the competitive solution. In *The collected works of F. A. Hayek: socialism and war*, ed. Bruce Caldwell. Indianapolis: Liberty Fund, 117-140.
- Hayek, F. A. 2010 [1952]. Scientism and the study of society. In *The collected works of F. A. Hayek, Volume 13: studies on the abuse and decline of reason*, ed. Bruce Caldwell. London: Routledge, 75-166.
- Hodgson, Geoffrey M. 1999. *Economics and utopia: why the learning economy is not the end of history*. London: Routledge.
- Holmes, Douglas R. 2009. Economy of words. *Cultural Anthropology*, 24 (3): 381-419.

- Keynes, John Maynard. 1936. *The general theory of employment, interest and money*. London: Macmillan.
- Lane, David A., and Robert R. Maxfield. 2005. Ontological uncertainty and innovation. *Journal of Evolutionary Economics*, 15 (1): 3-50.
- MacKenzie, Donald. 2012. Knowledge production in financial markets: credit default swaps, the ABX and the subprime crisis. *Economy and Society*, 41 (3): 335-359.
- Orléan, André. 2012. Knowledge in finance: objective value versus convention. In *Handbook of knowledge and economics*, eds. Richard Arena, Agnès Festré, and Nathalie Lazaric. Cheltenham: Edward Elgar, 313-336.
- Power, Michael. 2007. *Organized uncertainty: designing a world of risk management*. Oxford: Oxford University Press.
- Shackle, G. L. S. 1979. *Imagination and the nature of choice*. Edinburgh: Edinburgh University Press.
- Shackle, G. L. S. 1992 [1972]. *Epistemics and economics: a critique of economic doctrines*. New Brunswick: Transaction Publishers.
- Shiller, Robert J. 2005. *Irrational exuberance*. 2nd ed. Princeton: Princeton University Press.
- Skidelsky, Robert. 2006. Hayek versus Keynes: the road to reconciliation. In *The Cambridge companion to Hayek*, ed. Edward Feser. Cambridge: Cambridge University Press, 82-110.
- Skidelsky, Robert. 2009. *Keynes: the return of the master*. London: Allen Lane.
- Smith, Vernon L. 2008. *Rationality in economics: constructivist and ecological forms*. New York: Cambridge University Press.
- Soros, George. 2010. *The Soros lectures*. New York: Public Affairs.
- Surowiecki, James. 2004. *The wisdom of crowds: why the many are smarter than the few*. London: Abacus.
- Stark, David. 2009. *The sense of dissonance: accounts of worth in economic life*. Princeton: Princeton University Press.
- Tuckett, David. 2011. *Minding the markets: an emotional finance view of financial instability*. London: Palgrave Macmillan.
- Westbrook, David A. 2004. *A city of gold: an apology for global capitalism in a time of discontent*. New York: Routledge.

Richard Bronk is a Visiting Fellow at the European Institute of the London School of Economics and Political Science, where he taught from 2000-2007. He is author of *Progress and the invisible hand: the philosophy and economics of human advance* (Little Brown, 1998), and *The romantic economist: imagination in economics* (Cambridge University Press, 2009). His approach to philosophy of economics is grounded in a history of ideas perspective and in his practical experience in financial markets and economic policy.
 Contact email: <r.bronk@lse.ac.uk>
 Website: <www2.lse.ac.uk/europeanInstitute/staff/visitingStaff/bronk/Home.aspx>

Review of Mary S. Morgan's *The world in the model: how economists work and think*. Cambridge University Press, 2012, 435 pp.

ROBERT SUGDEN

University of East Anglia

The world in the model is a distillation of fifteen years of concentrated work by a major philosopher and historian of economics. Mary Morgan's topic is how economists use models. Her aim is "to offer both a history of the naturalization of modelling in economics and a naturalized philosophy of science for economics". Recognising that "science is messy", she presents "a series of historical case studies through which the philosophical commentary runs". She describes this commentary as an account of modelling "as a way of doing science that has its own rationale just as do other modes of science" (pp. xv-xvi).

This is a long book, written with great clarity and rich in historical detail. The opening chapter, which sets out the agenda for the rest of the book, includes discussions of Quesnay's *Tableau Economique*, arguably the first economic model, and of Frisch's model of macroeconomic cycles. The following chapters link methodological themes with extended historical studies of modelling practice. These studies feature Ricardo's model of the distribution of the economic product, the Edgeworth Box, the concept of the rational economic man, the Newlyn-Phillips Machine, the IS-LM model, the supply and demand diagram, and the Prisoner's Dilemma.

As a practising economic modeller with only limited knowledge of the history of economics, I found the case studies fascinating and illuminating. I particularly enjoyed Morgan's accounts of David Ricardo and of the Newlyn-Phillips Machine.

In conventional histories of economic thought, Ricardo is often presented as the first formal economic theorist. His approach is seen as substituting dry but mathematically tractable abstractions in place of Adam Smith's wide-ranging interest in actual economic life. His style of reasoning—working through numerical examples rather than general qualitative relationships—is seen as clumsy, or perhaps as a necessary concession to his readers' lack of mathematical sophistication. Morgan

gives us a very different picture, in which Ricardo's theory of production and distribution becomes "Ricardo's model farm" (pp. 73-79). Ricardo was not just a successful City financier who turned to economics. Following a traditional English route to social status, he bought a country estate and took an interest in farming. Morgan shows that Ricardo's numerical examples of production processes were similar to contemporary reports of agricultural experiments by gentleman-farmers. She suggests that Ricardo's modelling, far from being dry and abstract, was a generalisation of agricultural experimentation; his modelling strategy was to represent a whole economy as a single farm.

The hydraulic model of a macroeconomic system, first constructed by Newlyn and Phillips at the University of Leeds in 1949, is justly famous. In economics, it is usually seen as a curiosity and as an imaginative teaching aid. In computing, it is seen as one of the first analogue computers. In philosophy of science, it is an example of how a model can have a physical rather than mathematical existence, while being built out of materials very different from those of the reality that it represents. Its history is a good story, and Morgan tells it well. She shows that, in building their Machine, Newlyn and Phillips had to solve fundamental problems in macroeconomic theory concerning the integration of stocks and flows. They were able to do this because, unlike other macroeconomists of the time, they were working in the domain of hydraulics rather than in that of two-dimensional diagrams of mathematical equations.

The philosophical commentary on these case studies comprises both a history of the role that models have in fact played in economics and a methodological account of the practice of modelling. In Morgan's history, models became central to the practice of economic science only from the 1930s. Classical economics was based on the notion of 'universal laws'. Two new "styles of reasoning" began to emerge in the late nineteenth century, both of which were associated with the use of mathematics—the "method of mathematical postulation and proof" and the method of modelling (p. 18). Morgan recognises that the dividing line between these two styles is not easy to draw, and one might disagree with some of her categorisations. (For example, she presents Walras as a practitioner of mathematical postulation and proof, but his representation of a whole economy as a single auction might be viewed as a masterpiece of economic modelling.) In any event, Morgan is in no doubt that the method of modelling has come to dominate theoretical

economics. Models and modelling “are endemic at every level” of economics (p. 2); they have become “the primary way of doing economic science” (p. 379). Modern economics “has become a social science largely dependent on small mathematical or diagrammatic models, each separately representing different bits of the economy and each treated largely independently of the others” (pp. 378-379).

Morgan’s account of what models are, and of what modelling is, starts from an attempt to distinguish models from other forms of representation. For Morgan, the essential distinction is that models are “working objects” (p. 380); they provide resources for reasoning. She treats the concept of *manipulation* as fundamental to modelling: “representations only become models when they have the resources for manipulation” (p. 27).

Morgan repeatedly declares that models have two uses: “Models are objects to enquire into and to enquire with: economists enquire into the world of the economic model, and use them to enquire with into the economic world that the model represents” (p. 217). However, as the title of her book hints, she has more to say about the first of these uses than about the second. Summarising how economists use models, she says: “they [economists] ask questions, use the resources of the model to demonstrate something, and tell stories in the process” (pp. 217-218). This combination of activities is treated as “a kind of experiment” (p. 31); the similarity between modelling and experimenting marks a major difference from the method of mathematical postulation and proof (p. 33).

Morgan fleshes out this account by presenting “reasoning with models” as involving four steps. Step 1 is to construct a model relevant for some problem of interest. Step 2 is to ask questions about “something in the model or in the world”. Step 3 is to “demonstrate the answer to the questions using the model’s resources”. This is where manipulation comes in: the answer is arrived at by manipulating the model. Step 4 is to add a “narrative” that “accompanies the demonstration to link the answers back to the questions and to their domains” (p. 225). The implication seems to be that if the questions are about the world the answers have to come through the narrative (since the model can demonstrate only what is true of itself).

‘Manipulation’ and ‘demonstration’ seem to be almost interchangeable concepts in Morgan’s account. She says that a model comes with associated “rules for reasoning”, which are constrained

by “the kind of stuff that the model is made from, or language it is written in, or the format it has”. These rules “effectively determine the economist’s valid manipulation or use of that model”. For mathematical models, these are typically rules of geometry or algebra; for the Newlyn-Phillips Machine, they are rules of hydraulics (p. 26).

Take the case of Frisch’s model. The version discussed by Morgan consists of three dynamic equations linking the rate of production of consumption goods, the rate of production of capital goods and the stock of capital goods. These equations can be solved to give the time paths of the three variables from any given initial values. The model is of economic interest because the equations make sense as representations of properties of a real economy that is not subject to any exogenous cyclical shocks, but the time paths of the solution are cyclical. If this modelling exercise is to fit into Morgan’s schema, it seems that we must say that ‘the model’ is the three equations, ‘the kind of stuff the model is made from’ is mathematics, and the ‘manipulation’ is finding the solution to the equations. But the distinction between the model and its manipulation seems unhelpful here. The equations *define* the time paths of the variables; the cyclical paths are just as much the model as the equations are. Once Frisch has written down the equations, the solution has only to be found.

This point can be made even more plainly in relation to the Newlyn-Phillips Machine. Here, presumably, ‘the model’ is the system of tanks, pipes, and valves. As Morgan says, “the circulation and manipulation of the flows of water representing the flows of money are governed by the hydraulics” (p. 227). But ‘the hydraulics’ are just properties of the physical system; they are not actions taken by the modeller. Morgan has to say that putting the “manipulable resources” of this model to work is nothing more than switching on the Machine (p. 226, 14n).

It is perhaps significant that several of Morgan’s leading examples of models do not define their own ‘solutions’ in the way that Frisch’s model does. Take the Edgeworth Box. As understood by most economists, this is a two-dimensional diagram that can describe endowments and preferences in a two-person exchange economy. It does not specify what those individuals actually do, or what counts as an equilibrium solution. To derive any conclusions about the pattern of trade in the model, one has to add additional assumptions, for example that trade takes place at market-clearing prices, or that ‘blocked’ trades do not take place. If one treats the Edgeworth Box

diagram as ‘the model’, the way it is used in economics fits Morgan’s schema; adding assumptions about individual behaviour or specifying a solution concept can then be characterised as ‘manipulation’. But my sense is that economic theorists would be more inclined to say that, until those components are added, the Edgeworth Box is not a fully-specified model; it is only an extremely useful modelling component.

By treating models as things to be manipulated, Morgan’s schema encourages us to overlook the sense in which, even from the viewpoint of its own designer, a model is autonomous. As a modeller, I think of myself as constructing something that has a life of its own: having constructed it, I sit back and watch what it does. Viewed in this way, what is remarkable about the world in Frisch’s model is that it is a self-contained economy that generates its own cycles, through the workings of the mechanisms that are described by the model’s equations. Having seen that Frisch’s model economy has endogenous cycles that are similar to those we observe in real economies, we come to think that the real cycles might be caused by mechanisms similar to those of the model. But to think in this way, we have to see the model cycles as the product of the workings of the model, not of the modeller’s manipulations.

Morgan’s emphasis on the manipulability of models may be connected with what, for me, is the main limitation of her analysis—the thinness of its account of how models are informative about the economic world that they represent. Although she repeatedly says that models are used to answer questions about that world, she is remarkably reluctant to explain how they do this. How models help us to understand the world is “a messier problem that I return to shortly” (p. 31), “a much more difficult problem” (p. 33), “a very tricky topic to which I return” (p. 237), “a particularly thorny problem” (p. 392), and so on. But, as far as I can see, she does not offer a solution to the problem, beyond such vague formulations as “they [economists] trust that their models will capture some of the heart of the matter” (p. 246) or

narratives operate as a cognitive bridge between the abstract and still relatively general model with its demonstrations and the much more detailed accounts of the concrete events of the real economic world. In making these correspondence links, narratives offer potential explanations for these real-world events (p. 244).

Morgan seems to want to say that the explanation of real-world phenomena is only a secondary concern of modelling. Recall that in her schema, any such explanatory claims are contained in the ‘narrative’ of step 4. Describing her experience of listening to economists presenting models, she says:

Narratives occurred primarily in economists’ explorations into the world of the model [...] Of only secondary importance was that they provided the format for making informal or casual inferences from the model experiments to the events of the economic world (pp. 361-362).

At best, it seems, models provide only *informal*, *casual*, or *potential* explanations; any confidence in these explanations is a matter of *trust*.

I have to admit that much of my experience of economists’ presentations is similar to Morgan’s. But if explanation really is only a secondary concern of modelling, what in economics serves as a *formal*, *rigorous* or *actual* explanation of the concrete events of the real world? If I have read her correctly, her answer is ‘econometrics’.

Almost at the end of the book, she says:

No doubt, there were both descriptive and analytical aims for the early economists who created models, but there came a marked divergence after the interwar period so that statistical (econometric) modellers concentrated on theoretically informed descriptions that could be used for measurement and hypothesis testing, while mathematical modellers concentrated on providing accounts that established concepts and sparked hypothesis formation and theory development (p. 388).

It is not clear what Morgan means by “theory development” here, given her (I think correct) claim that modelling has become the primary way of doing economic science. If modelling is what economists mean by theory, how can there be a stage of theorising that progresses beyond modelling? It is easier to understand the claim that economics engages with the real world through econometrics. Presumably, econometrics is “theoretically informed” when it is informed by the kind of modelling that is the subject of Morgan’s book. But how do models provide *information* to econometricians, unless there is some reason to expect similarities between properties of models and properties of the real world? I cannot see how the practice of modelling can have scientific value unless models can provide genuine explanations of the phenomena of the real world.

Robert Sugden is professor of economics at the University of East Anglia (UK). His research uses a combination of theoretical, experimental, and philosophical methods to investigate issues in welfare economics, social choice, choice under uncertainty, the foundations of decision and game theory, the methodology of economics, and the evolution of social conventions. He is the author or editor of eight books, including *The economics of rights, co-operation, and welfare* (2005 [1986]) and *Experimental economics: rethinking the rules* (2009, with five co-authors). He is currently working on a book on contractarian welfare economics.

Contact e-mail: <R.Sugden@uea.ac.uk>

Review of Julian Reiss's *Philosophy of economics: a contemporary introduction*. Routledge, 2013, 352 pp.

HAROLD KINCAID

University of Cape Town

Writing a textbook on philosophy of economics presents some hard challenges. Neither philosophy nor economics is sufficiently defined and circumscribed to make an overview easy. The nature of philosophy is currently quite contested. The naturalism movement exemplified by Quine now competes with a very traditional, conceptual-analysis trend that gives limited place to science (and thus by implication to economics). Philosophy of science—which is at the heart of naturalism—has mostly moved away from broad issues such as the nature of confirmation or explanation to a close engagement with specific sciences, in large part because it seems not all that much can be said about explanation, confirmation and the like in the abstract. Economics has grown tremendously in terms of tools and topics. It is no longer possible (if it ever was) to identify the core of economics with static general equilibrium theory. Game theory, new institutional economics, and experimental economics—all applied to an increasingly wide range of phenomena—make it nearly impossible to talk about philosophical issues in the field as a whole.

Aside from these problems, a textbook in philosophy of economics confronts a problem of audience. Is the target philosophy students? If it is, then how is philosophy of economics supposed to be done for students without much knowledge of economics, especially if the goal is to do philosophy of science close to the science itself? If the target is economics students (though I doubt that more than a handful of economics departments offer such a course—mine does not), what distinguishes the course from an ordinary economics course on whatever part of economics is considered?

Julian Reiss's new book reflects these inevitable tensions but largely does an admirable job of confronting them. Reiss deals with the ever larger scope and available tools of economics and with the philosophy of science imperative to engage closely with the science by covering an enormous range of work in economics in a fair amount of detail, as well

as recent relevant general philosophy of science and normative theory. His general strategy is to explain some piece of economics in some detail, some relevant philosophy of science or ethical and political theory if needed, and then present something of a critical analysis of the economics and/or the philosophy. The relative emphases on these components vary from chapter to chapter. For example, some chapters are content to present the economics and the issues raised while others argue in detail for specific conclusions.

An enormous range of topics are included in this book, with a full one third of the book covering ethical, social, and political issues. Among the chapters are:

- A thorough presentation of standard decision theory;
- A presentation of rational-choice game theory;
- A survey of current views on causality and mechanisms in philosophy;
- A presentation of different approaches to idealized models;
- A discussion of the complexities of measurement in economics;
- An interesting discussion of present day emphasis in economics on randomized controlled trials and natural experiments;
- A catalogue of uses that economic experiments might serve and the kinds of difficulties they face;
- A discussion of accounts of well-being and distributive justice;
- Arguments for libertarian or “soft” paternalism.

The end result is that Reiss’s book allows for tremendous flexibility in designing different courses. For example, it would be easy to develop courses that involved nearly all philosophy of science issues or ones that were mostly normative issues around economics, with any mix in between. I also think Reiss’s text can work as the sole book for a course or be used along with other articles to expand on the discussion Reiss provides. The book certainly provides much material that can usefully be expanded upon in lectures. There is nothing out there that rivals it for teaching a course on philosophy of economics, and it would be a good tool for certain kinds of courses in political philosophy or philosophy and public policy.

While Reiss clearly intends his work to be a textbook in the field (there are useful study questions and chapter summaries), he cannot help making some substantive claims in philosophy of economics.

Many of these are compelling. For instance, Reiss argues convincingly that normative values are involved in fundamental ways in making basic economic measurements. He works through the details of various ways of constructing the consumer price index, for example, and identifies the value assumptions involved. His conclusion is not that such measurements are worthless but rather that the particular value assumptions involved should be made explicit.

This is the kind of philosophy of economics I like. It is close to the details and recognizes the variation on the ground—the different ways theories, models, and so forth, are applied. It does not rely on traditional philosophical methods of conceptual analysis with necessary and sufficient conditions tested against what we would say or our intuitions. It does use the traditional philosophical skills of careful separation of theses and arguments to clarify the science and perhaps to contribute to its betterment.

Where I disagree with Reiss's substantive claims in philosophy of economics is when he drops this approach and returns to traditional philosophical methods of conceptual analysis. He does so on several occasions. I will discuss two: his treatment of rational choice theory and of game theory.

Reiss takes microeconomics to be based on and to be defending a normative account of rationality. He also takes the less ambitious approach of revealed preference theory as an account of the notions of choice and preference, where 'account' here means supplying individually necessary and jointly sufficient conditions for fundamental concepts. Assuming that economics is committed to such goals, he is then able to identify counterexamples where economists' accounts do not capture standard intuitions about rationality and choice.

Reiss is not unusual in looking at microeconomics in this way (see Hausman 2012), and maybe some economists do so as well. However, there are other ways to understand what economists do that do not commit them to the project of specifying a general theory of rationality or to take revealed preference theory as defining choice and preference. Good science often develops concepts that have no clear connection or basis to those of common sense and trying to force science to have those connections can be a positive hindrance to scientific advancement (see Ross, et al. 2013). There is a plausible history of economic thinking that sees developments in the 20th century as moving farther and farther away from ordinary common

sense and folk psychological notions of choice and rationality towards quite distinct notions that suit economists' purposes.

Those purposes have been above all to explain aggregate market supply and demand phenomena (see Kincaid 1996). Concepts of rationality and choice have thus been moulded to that end. Rather than trying to provide a general theory or a definition of rationality or rational behaviour—which Ken Binmore (2011) points out is like trying to define life or conscious, which is unlikely to be helpful—the concern is to find constraints suitable for modelling aggregate behaviour. The standard requirements are made to identify consistent patterns of choice. It is an empirical question whether such patterns exist in aggregate market behaviour. Whether 'consistent patterns of choice' captures everything in common sense notions of rationality is beside the point. Something similar is true of revealed preference theory. By design it is not about the underlying processes behind choice, nor is it a theory of choice in the folk psychological sense. Revealed preference theory is about consistent behaviours that are sensitive to incentives. Behaviours combined with prices and budget constraints are designed to help understand aggregate market phenomena. Once again the kinds of counterexamples that Reiss cites showing that behaviours of this sort do not match our common sense notions of choice and preference are beside the point.

I am sure that identifying exactly what economists are up to with revealed preference theory is not an easy task. As mentioned above in my positive remarks about Reiss's discussion of the consumer price index, economists may invoke revealed preference theory for different purposes and with different understandings in different contexts. Those have to be sorted out with care, the clearest attempt to do so on my view being Don Ross's (2011; 2012). Such sorting out requires precisely *not* confusing revealed preference theory with folk psychological notions.

A second place where Reiss's philosophy of economics is too traditional for me is in his discussion of game theory. The following quotation summarizes his view nicely:

Game theory, understood as a theory of rational decision-making, is thus highly problematic. The Nash equilibrium is ill-justified. Even if it were justified, it would solve few problems because most games have multiple Nash equilibria. Thus far, the refinement

program has produced few results that can be defended from the point of view of rationality (p. 73).

What bothers me about this approach is that it dismisses a large body of empirical social science on general philosophical grounds, namely, that game theory cannot be defended as an instantiation of general principles of rationality.

There are many possible responses to Reiss's scepticism. First, it completely ignores evolutionary game theory—it is not even mentioned in the book—despite the fact that it has probably eclipsed rational-choice game theory in economic applications. Evolutionary game theory obviously does not depend on a theory of rationality for its results and thus is not subject to Reiss's criticisms even if those criticisms are compelling for rational-choice game theory.

Moreover, Reiss's criticisms of rational-choice game theory are not convincing when it comes to uses of game theory *as empirical science* (whether they are compelling criticisms of the uses philosophers make of game theory is another issue). The goal in empirical applications is to use game theory as a modelling tool, not to develop an a-priori theory of rationality. Reiss claims that Nash equilibrium is unjustified. However, past supposed counterexamples to Nash equilibrium in the literature have been often shown to be implicitly invoking a different game than the one explicitly represented. Ken Binmore (1994) has shown this for standard alleged counterexamples. However, even if this were not the case, there are many applications of game theory where we know Nash outcomes result, regardless of whether being a Nash equilibrium is a necessary and sufficient condition of being rational. Of course evolutionary game theory is such a case, for evolutionary stable strategies are Nash outcomes. Furthermore, when given sufficient incentives and sufficient time to learn, experimental subjects find Nash outcomes even in bargaining games, widely thought to be the most difficult domain for successful game theory predictions (see Binmore 2007). They also do so when there are multiple Nash equilibria.

There are of course refinements to Nash equilibria and alternative equilibrium concepts. For example, one particularly important alternative is the notion of quantal response equilibria (QREs) which allow for equilibria around Nash points and which leave room for one sense of errors. In empirical work QREs allow for predictive success that a simple Nash equilibrium does not. So it is not a matter of which is the "right" equilibrium. Rather it is an empirical and modelling question

about which notions fit the data best; there will not be one answer across all contexts.

I think there is no doubt that sometimes game theory works. Industrial organization, a vibrant area of microeconomics, is now almost entirely done using game theory tools. I would not want to dismiss that work *carte blanche* because of philosophical doubts that Nash equilibrium is rational. I would want to know how that work deals with problems of multiple equilibria, what solution concepts it uses and why, and so on. There is of course no guarantee that all applications are successful or believable. Yet that is a judgment that has to be made while looking at the full gory details of the applications. The philosophy of economics has to be closer to the economics itself than Reiss's discussion provides.

However, while I have some philosophical differences with the author, that adds rather than detracts from the book's value as a textbook. Reiss clearly takes stands on issues. Doing so makes for a much better textbook over one that just blandly rehearses different positions in debates. It is good that there are issues to argue with in his textbook.

REFERENCES

- Binmore, Ken. 1994. *Game theory and the social contract*. Cambridge: MIT Press.
- Binmore, Ken. 2007. *Does game theory work?: the bargaining challenge*. Cambridge: MIT Press.
- Binmore, Ken. 2011. *Rational decisions*. Princeton: Princeton University Press.
- Hausman, Daniel M. 2012. *Preference, value, choice, and welfare*. Cambridge: Cambridge University Press.
- Kincaid, Harold. 1996. *Philosophical foundations of the social sciences*. Cambridge: Cambridge University Press.
- Ross, Don. 2011. Estranged parents and a schizophrenic child: choice in economics, psychology, and neuroeconomics. *Journal of Economic Methodology*, 18 (3): 215-229.
- Ross, Don. 2012. The economic agent: not human, but important. In *Handbook of the philosophy of science, volume 13: philosophy of economics*, ed. Uskali Mäki. Amsterdam: Elsevier, 691-735.
- Ross, Don, James Ladyman, and Harold Kincaid (eds.). 2013. *Scientific metaphysics*. Oxford: Oxford University Press.

Harold Kincaid is professor of economics at the University of Cape Town and director of the Research Unit for Behavioural Economics and Neuroeconomics (RUBEN). He is co-editor of *The Oxford handbook of*

philosophy of economics (2009) and editor of *The Oxford handbook of philosophy of social science* (2012).

Contact e-mail: <harold.kincaid@uct.ac.za>

Review of Wilfred Beckerman's *Economics as applied ethics: value judgements in welfare economics*. London: Palgrave Macmillan, 2011, 240 pp.

JONATHAN B. WIGHT
University of Richmond

This is a well-written textbook geared to advanced undergraduate or graduate students of economics, many of whom are largely and regrettably innocent of the ethical problems inherent in conventional economic analysis. It compares with Daniel M. Hausman and Michael S. McPherson's *Economic analysis, moral philosophy, and public policy* (2006) and Johan J. Graafland's *Economics, ethics, and the market: introduction and applications* (2007). The book presupposes a fair amount of knowledge in both economics and ethics (it does not intend to be a primer in either). The author is professor emeritus at Balliol College, Oxford and honorary visiting professor of economics at University College London, and this book arose from a third-year course at the latter school in which he participated.

The book contains 17 chapters, each of them reasonably concise (ranging from 8 pages to 24 pages). Each chapter could thus be covered on a single day or two in class. The topics are generally what one would expect: the fact-value dichotomy; getting from individual choices to individual welfare and from individual welfare to social welfare; utilitarianism and its critics; GDP; happiness; equality; justice; the value of life; and the bounds of moral standing in space and time (e.g., international and inter-generational welfare).

The author is dissatisfied with the standard way of teaching economics, which supposes that one can do policy making without carefully addressing ethical precepts. Another frustration with current economic teaching is the focus on optimality rather than on understanding our actual, second-best, starting position. The book attempts to demonstrate, in chapter 1, the mix of value judgments with facts necessary for welfare economics. Two examples are provided, one of natural resource depletion and the other of global warming. Citing Ian Little, Beckerman notes that in both cases finding a unique optimum solution is impossible given that ethical choices dominate

any conclusion. Hence, “‘Optimal’ does not mean ‘ethical’” (p. 11). In addition, I would add that historical, cultural, and political frameworks create institutional regimes with path dependencies that are important for policy analysis and that go far beyond the scope of efficiency analysis. Dani Rodrik’s *One economics, many recipes: globalization, institutions, and economic growth* (2007) is suggestive of what can and should be done to incorporate some of these elements.

Beckerman does not do all that he could with the fact-value dichotomy. For example, he accepts that scientific propositions in positive economics depend upon factual observations. But an unexamined issue in the book is how values are used to create the very facts that are presented as scientific. For example, the “fact” that the unemployment rate is 7% depends in part on the value judgments that go into the official definition of who is “unemployed”. A looser definition could produce policies that result in more fiscal stimulus and more people on welfare rolls. In economics, definitional terms are socially created and they reflect professional judgments mixed with moral and political norms.

Chapter 3 discusses the trade-offs between policy goals, such as low inflation and low unemployment. The point is to demonstrate that “both value judgments and positive propositions must enter, sooner or later, into any specific normative economics prescription” (p. 33). In many instances the ethical judgments in welfare economics are “not always adequately appreciated” (p. 33)—which is likely an understatement. In my own experience, colleagues have argued that once a value proposition is widely adopted it is no longer “ethical” it is objectively “professional”. This is misguided. For example, a professional norm in econometrics is to accept a Type I error rate of 5 percent, i.e., which produces false positives in 5 out of 100 cases. However, the loss function for a Type I error varies greatly depending on the circumstances. If a food additive is suspected of causing serious brain injury in children, would one prefer to make fewer false positives in this case, compared to when the loss involves merely an upset stomach? There is no objective answer to this question because the choice requires a value judgment. Establishing a professional norm cannot diminish the ethical significance of what is at stake.

Using David Hume’s and Adam Smith’s writings, Beckerman adeptly demonstrates that ethical analysis in economics has deep historical roots. In introducing altruism and benevolent sympathy he cites the

familiar opening sentence from Adam Smith's *The theory of moral sentiments* (1759):

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it (Smith TMS, I.i.1).

Yet readers should be aware that Smith was also interested in anti-social instincts for hatred and revenge that play an important part in the development of institutions of justice.

In chapter 4, on social welfare functions, Beckerman laments that standard microeconomics insists that no interpersonal comparisons of utility can be valid. While the neoclassical approach is scientifically logical, it defies our moral intuitions: a loaf of bread taken from a plump person and given to a starving child would certainly increase human welfare in the minds of most people, even if there is no way of proving this. Given the growing interplay between economics and biology, there may indeed be ways to calculate substantive measures (e.g., hormonal responses) of well-being in the future. Beckerman notes, using the wry humour that peppers the book:

We all know that to rule out interpersonal comparisons of this kind simply because there is no scientific basis for them is nonsense. Rigour is extremely important, but it must not be allowed to become 'rigor mortis'. If we were unable to empathize with other people sufficiently for us to be able to make fairly sound judgements about significant differences in levels of utility, the social concepts of fairness that are essential in any stable society would be impossible (p. 62).

To Beckerman, making "rough interpersonal comparisons" between utilities can be justified in specific circumstances.

Beckerman addresses the limitations of Larry Summers's famous memorandum on exporting pollution to impoverished African countries.¹ He rightly points out the distributional as well as principal-agent problems with Summers's proposal, and the fact that compensation for pollution victims is unlikely. But he might also have pointed out that there are few institutional safeguards in the countries

¹ This was part of an internal memo on trade liberalization circulated in December 1991, while Summers was chief economist at the World Bank, and which was published in *The Economist*, 8th February 1992.

concerned, whether for legal redress or political protest. In other words, Summers (like most economic analysts) assumes the existence of functioning institutions that are essential for justifying the implicit coercion that lies behind modern cost-benefit calculations. The Chinese government forcibly relocated 1 million peasants to build the Three Gorges Dam based on purported economic benefits and costs. But opportunity costs cannot be calculated in involuntary transactions, and without judicial restraints, a free press, and a democratic process, these estimates are not subject to impartial checks. In a Hobbesian world, coercion is justified only when the contractual basis for coercion is voluntarily agreed to ahead of time.

Ultimately, Beckerman is sanguine about the role of cost-benefit analysis in the economic toolkit:

It might appear from the above that welfare economics cannot take us very far in making rational choices among different policy options. But this would be a mistake. Welfare economics, combined with the social welfare function, provides a valuable framework and organizing principle for taking account of the effect of any economic policy (p. 76).

Because most policy choices are about marginal adjustments to resource allocations, “what is usually required will be factual information, and there will be little point in wringing one’s hands over the normative significance of the starting point” (pp. 76-77). Nevertheless, a sterile cost-benefit analysis that fails to adequately address ethical values may become a harmful activity.

Overall, this book is highly recommended. It covers the selected topics with depth and sensitivity. The writing is generally excellent, but there are occasions of repetition and unevenness, as if the chapters were compiled separately and merged later. A student reader who is not already familiar with basic ethical theories could benefit from a primer in some places. For example, the book discusses Amartya Sen’s theory of commitment, however it does not dig very deeply to explain or defend that notion, whether from a deontological or virtue ethics approach.

The book devotes a lot of attention to questions of equality and justice, particularly on the work of economist philosophers such as John Broome, Partha Dasgupta, Ian Little, and Amartya Sen. This is appropriate, interesting, and relevant. However, the book does not

appear to address research in experimental economics, biology, and psychology that might be relevant to some of these questions, such as the work in neuroeconomics by Paul Zak, experimental work by Vernon Smith, or recent philosophical work on virtue ethics by Deirdre McCloskey. This is the normal limitation of any text that strives to be concise, yet students should understand there is much more to ethics and economics than can be conveyed in this book.

In particular, if economics is applied ethics, as the title suggests, economists themselves must be ethical in the pursuit of science. However, the book does not address the moral responsibilities of being a scientist, nor does it address the conflicts of interest of economists before and during the crisis of 2008 (as alleged in the documentary, *Inside job*). These would seem to be important topics for economics students but they lie beyond the scope of this work. As with the ethics embedded in normative economics, the ethics embedded in positive economics also remain generally unexamined.

REFERENCES

- Graafland, Johan J. 2007. *Economics, ethics, and the market: introduction and applications*. New York: Routledge.
- Hausman, Daniel M., and Michael S. McPherson. 2006 [1996]. *Economic analysis, moral philosophy, and public policy*. Cambridge: Cambridge University Press.
- Rodrik, Dani. 2007. *One economics, many recipes: globalization, institutions, and economic growth*. Princeton: Princeton University Press.
- Smith, Adam. 1790 [1759]. *The theory of moral sentiments* [TMS]. London: A. Millar.
<http://www.econlib.org/library/Smith/smMS.html> (accessed April 2013).

Jonathan B. Wight is professor of economics at the University of Richmond. He is the author of the academic novel, *Saving Adam Smith: a tale of wealth, transformation, and virtue* (2002); and a book of lessons on ethics and economics for teachers, *Teaching the ethical foundations of economics* (2007, co-edited with John S. Morton). He currently serves as president of the Association for Social Economics.
 Contact e-mail: <jwight@richmond.edu>

Review of Andrew M. Yuengert's *Approximating prudence: Aristotelian practical wisdom and economic models of choice*. Palgrave Macmillan, 2012, 246 pp.

RICARDO F. CRESPO

IAE (Universidad Austral) and Universidad Nacional de Cuyo

Although it may sound paradoxical, this is a *positive* book about the *limitations* of economics. All sciences necessarily simplify. Sciences try to think deeply about their subjects, and to think we need to put away the details and concentrate on the essence of our subject. However, we do not tend to think about what we have put away. This is important because it can happen that, forced by the requirements of tractability, we put away relevant 'details'. Yuengert shows in this book that economic modeling undertakes only a partial analysis of economic action, because it 'puts away' interesting features of its subject that deserve to be taken into account. He proposes adopting the Aristotelian account of human action—more specifically, of practical wisdom—as the benchmark against which to consider economic modeling. He maintains that “economics can learn much about its limits from Aristotle, who describes aspects of choice behavior that cannot be precisely modeled” (p. 3). Thus, the aim of the book is to determine what aspects of human behavior cannot be captured by the economists' models. In this task, Yuengert has the advantage of being a well-informed and up to date academic economist: an economist talking to economists. He knows the current literature on economics' new perspectives, from behavioral economics to neuro-economics to economic sociology. And he provides technical examples familiar to economists. Yuengert has also has the advantage of having studied philosophy with the aim of enlightening economics. Thus he is able in this book to present philosophical concepts and arguments in a way that economists can appreciate.

Chapter 1 introduces the book. Chapter 2 justifies the need to compare economic modeling with the Aristotelian philosophy of action. The fact that economists consider that models are approximations implies that there is a difference between models and a 'background account' of actual choice. This chapter carefully analyses why the

traditional economic optimization model is insufficient to solve complicated decision problems in the face of pervasive uncertainty, and argues for the need to incorporate a form of reasoning other than the instrumental. We need, Yuengert maintains, a comprehensive view of human choice acknowledging—as the economists claim—that economic agents act intentionally, for reasons. Yuengert takes reasons as causes in the context of a free human agent. It also presumes that human rational decisions are not only calculative, that instrumental rationality does not exhaust rationality. Such concerns lead directly to the Aristotelian theory of human action.

Chapter 3 introduces the Aristotelian concept of practical wisdom (or ‘prudence’). This human capacity integrates human reason, emotion, habit, and instinctive traits, to make decisions in a non-deterministic way. Yuengert develops the differences between practical wisdom and technique. While practical wisdom deals with both means and ends, technique only deals with the means to attain *given* ends. That is, for technique ends are fixed and external, while for practical wisdom they are dynamic and internal. The differences between practical wisdom and the logic of constrained optimization—technique—can already be discerned, and they are developed in the following chapters.

Chapter 4 makes a comparison between well-behaved objective economic functions and the Aristotelian explanation of action. One problem with the economic concept of the utility function is that in real life preferences are not given—they are dynamic and discovered in the very process of acting. We do not start off with a set of well-behaved preferences and then act; means and ends are mutually determining each other *while* we act. Yuengert carefully shows how this complex interaction of means and ends cannot be grasped by cost-benefit analysis techniques. Another problem is the incomparability of our ultimate ends, which leads to incompleteness in preference orderings. The homogenous concept of utility is inadequate for such cases.

A minor point to note in chapter 4 is that Yuengert adopts a specific interpretation of Aristotle’s concept of *eudaimonia* (a word often defectively translated as ‘happiness’). There are two main interpretations of the meaning of Aristotle’s *eudaimonia*. The ‘inclusive view’ of *eudaimonia* promoted by John Lloyd Ackrill (1980) and adopted by Yuengert holds that *eudaimonia* is an inclusive end composed or

constituted out of defined “second order” ends.¹ It is more practically oriented. The other interpretation, by Richard Kraut (1989), maintains that *eudaimonia* is a dominant end different from second order ends. Second order ends are sought not only for the sake of themselves but also always for the sake of the *eudaimonia* to which they are subordinated, but for which they are not always necessary. This view is more oriented to theoretical contemplation than practical action. For example, a sick person might also be *eudaimon* in this account: it all depends on how she copes with her illness.

Aristotle leaves room for both interpretations. In effect, in his *Nicomachean ethics* he develops an account of the virtues leading to *eudaimonia*. But at the end of the *Ethics* he makes clear that a ‘perfect’ (*teleia*) *eudaimonia* is a contemplative or theoretical activity (*Nicomachean ethics* X, 7, and 8). However, these two versions of *eudaimonia* are compatible, especially when considering them in the wider context of Aristotle’s thought: there are some people called to practical life and others called to theoretical life, and there might also be different stages in life which call for one or the other. Both virtues and material goods are necessary for both kinds of *eudaimonic* lives. Yuengert, as I mentioned, declares that he adopts Ackrill’s position, but in fact he also seems to follow Kraut’s when he asserts that *eudaimonia* is not contained in the second-order ends, but in their ordering (p. 53).

Chapter 5 outlines the difference between the Aristotelian concept of contingency—the singularity of human actions—and the economic concept of uncertainty. Much depends here on what concept of uncertainty economics adopts. Mathematicians and economists have generally tried to avoid contemplating the contingency of future events. Only a few economists have warned about the unavoidability of contingency in many fields: Frank Knight, John Maynard Keynes, George L. S. Shackle, and Friedrich von Hayek. In 1921, Knight distinguished risk—where there is an objective probability and it is known—from subjective probability—where there “is no valid basis of any kind for classifying instances” (Knight 1921, 225). Keynes expressed

¹ We can distinguish between a) ends that can be considered only as means, only pursued for the sake of something else (first-order or instrumental ends), b) ends that are desirable in themselves and also pursued for the sake of some other final end (second-order ends), and c) ends which are only desirable in themselves (third-order or final ends: usually known as *eudaimonia* or ‘happiness’). For example, we study for an exam (i.e., a means to a first order instrumental end) in order to achieve graduation (a second-order end), in order to be happy (a final end) according to our plan of life (designed by practical reason).

this in a very similar way in his famous 1937 paper: “about these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know” (Keynes 1937, 113).

Yet, despite the inapplicability of probability calculations to such matters, people need to act, and they do use probability estimates in deciding what to do. Though this appears rigorous, it is not. This process has nevertheless acquired a scientific character thanks to the mathematical talents of people like Frank Ramsey, Bruno de Finetti, and especially Leonard Savage. Savage (1954) argued that people behave *as if* they have a subjective *a priori* belief about the probability of future events, and that this can be discovered through the empirical examination of people’s decisions *a posteriori*. This move distorted Knight’s concept of uncertainty. What was purely subjective for Knight, or uncertain for Keynes, became ‘objective’ for Savage (he called it the “personalistic theory of decision”). Savage’s proposal became expected utility theory, the dominant paradigm of contemporary economic decision theory.

Yuengert considers Knightian uncertainty to be compatible with the Aristotelian concept of contingency. However, he does not take into account one possibility considered by Aristotle that would partially overcome the extreme difficulty of dealing scientifically with singular facts. Aristotle distinguishes between three classes of facts: necessary facts which always occur in the same way; general facts which mostly occur in the same way; and accidental facts which scarcely ever occur in the same way (*Physics* II, 5, 196b 10ff. and *Metaphysics* VI, 2, 1026b 27ff.). The exact sciences deal with the first category, physics and practical sciences with the second, and the third cannot be the subject-matter of any science. “General facts” are *hos epi to polu* (those which occur in many cases, but not of necessity or always). This is an expression not only used in the quoted passages from the *Metaphysics* and *Physics*, but also in the *Nicomachean ethics* (I, 2, 1094b 21), in reference to the practical realm. Given that, by definition, statistics deals with general facts, it is clear that it cannot be an exact science in Aristotle’s sense. This does not imply its weakness, but rather the need to adjust our expectations of it to the nature of its subject-matter. Nevertheless, this inexactness is of a different kind to the inexactness that comes with dealing with singular cases.

This leads me to note that this book does not consider Aristotle’s account of practical science, which is different from his practical

wisdom.² Aristotle thought that there were some regularities in the human realm that could be the object of a science, though an inexact one. In fact, Yuengert implicitly refers to practical science, for example when he praises case-based decision theory (e.g., Gilboa and Schmeidler 2001), in which economic agents cope with contingency by relying on memory and looking for similarities between cases. However, if Yuengert had explicitly considered Aristotle's practical science, he would have found fewer difficulties with economics. As Aristotle's ethics and politics demonstrate, it is possible to build an inexact science of tendencies. It is true, as Yuengert explains excellently, that Aristotle's remedy for contingency is based on the virtues. However, I think that this does not exclude the possibility of a positive account of generally repeated conduct.

Chapter 6 returns to virtue. After explaining the Aristotelian concept and characterization of virtues, Yuengert presents four types of human conduct: 1. the virtuous, 2. the continent, 3. the incontinent and 4. the vicious. If we only observe the external outcomes, the first two (1 and 2) and the last two (3 and 4) are indistinguishable, but in 2 and 3 there is an internal conflict. Virtue helps to overcome the conflicts of the continent and incontinent actors. Yuengert notes that recent economics research into addiction, behavioral economics, human capital and time inconsistency models try to take this internal conflict into account, and he explains why they not fully achieve that. While the synthetic character of practical wisdom can address the problems identified in this and the preceding chapters, Yuengert concludes in chapter 7 that economics cannot attain that level of comprehensiveness.

Chapter 8 describes the characteristics of the *phronimos*, the prudent person. There is no formula or manual containing the rules of practical wisdom, there are only practically wise persons. Yuengert finds in the literature on tacit knowledge (e.g., Polanyi 1966), learning by doing and social norms some recognition of the traits of the *phronimos*, without being capable of modeling him.

² We should distinguish practical science, practical reason, and practical wisdom. Practical wisdom is the virtue of prudence, correctly characterized by Yuengert. Practical reason is a discursive form of thinking about what we should do: in deliberating about our purposes or ends, it produces a judgment that is used by practical wisdom. Practical philosophy or science is both a discipline and a critical reflection on practical reasoning, its process, and its goals. It deals with those subjects relating to human decision or choice and it has a practical aim (*Metaphysics* II, 1, 993b, 21-22; see also *Nicomachean ethics* I, 2, 1095a 6, and II, 2, 1103b 27-28).

Chapter 9 is the final chapter. Given that the book has shown that “any quantitative optimization model of human decision making cannot hope to be comprehensive” (p. 158), this chapter argues that economists must be mindful of the limits of their models, that “economists ought to be mindful that the economic method is a small worlds approach, which consequently cannot speak comprehensively to a large worlds reality” (pp. 160-161). The limited conclusions permitted by the idealizations built into economic models might be enough for positive analysis, but are not sufficient for normative analysis. The details left out of the models are necessary for normative work. True normative economics, Yuengert concludes, requires practical wisdom.

The book ends with two appendices: the first on the need for and meaning of realism in economics, and the second about naturalistic and non-naturalistic accounts of social reality.

This book provides the useful service of identifying the characteristics of human action that economic models cannot take into account. It is useful because it explains the challenge to positive economists of trying to incorporate these characteristics into their approach, and because it highlights the features that economists must consider in their normative work. The contribution of the book lies in its originality. Economics books are not usually about what economics cannot do.

REFERENCES

- Aristotle. 1995. *The complete works of Aristotle: the revised Oxford translation*, ed. Jonathan Barnes. Princeton: Princeton University Press.
- Akrill, John Lloyd. 1980. Aristotle on eudaimonia. In *Essays on Aristotle's ethics*, ed. Amélie Rorty. Berkeley: University of California Press.
- Gilboa, Itzhak, and David Schmeidler. 2001. *A theory of case-based decisions*. Cambridge: Cambridge University Press.
- Keynes, John Maynard. 1937. The general theory of employment. *The Quarterly Journal of Economics*, 51 (2): 209-233.
- Knight, Frank H. 1921. *Risk, uncertainty, and profit*. Boston: Houghton Mifflin.
- Kraut, Richard. 1989. *Aristotle on the human good*. Princeton: Princeton University Press.
- Polanyi, Michael. 1958. *The tacit dimension*. New York: Doubleday.
- Savage, Leonard J. 1954. *The foundations of statistics*. New York: Wiley.

Ricardo F. Crespo holds BA degrees in economics and in philosophy, as well as doctorates in philosophy (Universidad Nacional de Cuyo, Mendoza, Argentina) and in economics (Universiteit van Amsterdam).

He teaches philosophy of economics at IAE Business School (Universidad Austral) and Universidad Nacional de Cuyo and he is a researcher at Argentina's National Council of Scientific Research (CONICET). His recent publications include *Theoretical and practical reason in economics: capacities and capabilities* (Springer, 2013) and *A re-assessment of Aristotle's economic thought* (Routledge, forthcoming). Contact e-mail: <RCrespo@iae.edu.ar>

Review of James Halteman and Edd S. Noell's *Reckoning with markets: moral reflection in economics*. Oxford University Press, 2012, 218 pp.

BETSY JANE CLARY
College of Charleston

Halteman and Noell have written a timely book on the role of ethics, values, and value judgments in economic discourse. The recent financial crisis has brought to light many very objectionable business practices, all pursued in the name of money and often at great cost even to those on whose behalf the financial institutions were ostensibly acting. It is an indictment of the current state of economic theory and practice that many of those who instigated these practices never had any idea that they were engaging in unethical behavior. This book shows how economics as a discipline has failed to provide the necessary moral framework by which economic behavior can be judged. This is a shame, and it has not always been the case, for much in the history of economic theory is available to guide moral economic conduct. Halteman and Noell trace the history and development of moral reflection in economic theory, including those instances in its history when economic theory has been devoid of moral considerations.

The book is written for the undergraduate student and would be useful in courses on the history of economic thought, behavioral economics, economic psychology, and business ethics, among others. The authors include illustrations applying the various ethical perspectives of past economists to current economic practices and policies. A set of discussion questions is included with each illustration.

The authors begin the discussion of moral reflection with Plato and Aristotle and the concept of the human *telos*. Halteman and Noell argue that for Plato and Aristotle, as well as for others, the concept of *telos* involved some idea of what people were meant to be. Thus, these philosophers believed that as people struggled with the conflicts between human nature and their ultimate purpose in life, rules of moral conduct, or ethics, developed to maintain order in society. Halteman and Noell claim that in classical Greek thought behavior was measured in terms of right and good conduct; that equitable economic distribution

was related to the requirements of one's social station; and that human flourishing within the community was the criteria by which behavior was judged. Thus, in contrast to much modern economic thought, the group, or society, superseded the individual in ancient thought. The formation of a social order was not understood as a natural process. It was a product of dialogue and moral reflection about justice which led to institutionalized laws, rules, and norms for determining and administering virtuous (right) behavior among the individuals in society.

The authors include examples of economic situations in which the understanding of virtue guides behavior in achieving human flourishing. In Plato, the solution to the problem of scarcity is for people to desire less and recognize the virtue of a simple life. The authors point out that the ancient Greeks and Hebrews, and the Christian Church throughout the Middle Ages, all focused on the problem of human desires rather than on the economic problem of production. As an example of Aristotle's notion of teleology as the guiding principle of economic life, the authors use his example of a shoe, which has the natural purpose of helping one to walk and the unnatural purpose of being an object of speculation for monetary gain. Exchanging shoes for their use is therefore an appropriate reason for trade, while exchanging shoes for speculation is unnatural and morally wrong. Given that this book was published in 2012, a similar point could have been made using the example of a house.

The book continues with a discussion of Scholastic thought on economic justice, and the authors conclude this chapter with a very timely analysis of the relationship between Scholastic thinking on usury, avarice, and unjust gain and the subprime mortgage crisis. Yet, while the authors explain that the Scholastics recognized certain "exceptions to the usury prohibition", they neglect how far this went. The Scholastics made a clear distinction between usury and the legitimate payment of interest on a loan. Usury was the payment of unjustified interest, but not all interest was usurious. In Scholastic thought the payment of interest was justified as long as the interest was to cover some cost to the lender. The cost could be the opportunity cost of the profits not made by the lender's use of the funds, or the cost could be due to risk, either the risk of the venture funded by the loan or the risk of not being repaid. In either case, the loss to the lender had to be real in order for the payment of interest to be justified.

Halteman and Noell's claim that Calvin reversed Scholastic thought on usury and affirmed the legitimacy of charging interest on loans, except to the poor, also misses this distinction between usury and legitimate interest. Usury is still practiced, as evidenced by the sub-prime loan scandal. According to Robert Skidelsky, Maynard Keynes considered the rate of interest set by the market to be the most unjust price in the economic system and used the term "usury" to condemn it (Skidelsky 2009, 149). Nevertheless Halteman and Noell correctly point out that the value-free notion of the market in terms of supply-and-demand overlooks the rich understanding of the ethics of a modern financial system available in Scholastic thought. Their discussion of the Scholastics provides a substantive addition to this understanding.

The chapter on Adam Smith opens with an informative distinction between moral reflection related to a sense of purpose and the role of moral reflection in a mechanistic system. The authors point out that while moral reflection and dialogue are necessary to determining right and appropriate behavior for achieving one's goal(s), such a process of reflection, dialogue, and moral discernment is unnecessary in a mechanistic system which operates according to objective natural laws wholly independent of human interests and purposes. The question then becomes whether the economy is like a great machine, operating with no guiding purpose, or whether the economy does exist for some purpose and requires moral dialogue to function properly. The authors conclude, correctly, that Smith did not separate fact (positive economics) from value (normative economics); that for Smith moral reflection is an integral part of economic thinking; and that the purpose or goal of the economy for Smith was human flourishing and well-being for all.

One of the illustrations applying Smith to current practice concerns Smith's views toward market behavior and regulation. Halteman and Noell conclude that Smith did not appeal to the need for legislative intervention in the market. They argue, instead, that Smith disputed "the ability of governmental authorities to act in a manner consistent with justice to correct these cases of abuse by market combinations" (pp. 88-89). They base their conclusions on Ronald Coase, arguing that Smith would "presumably prefer that self-regulating social forces be given the opportunity to overcome deceitful commercial practices" (p. 89). However, Smith gave more importance to the role of government in enforcing justice than Halteman and Noell appear to allow. Where

social forces failed to overcome such practices, it was one of the duties of the legislator, i.e., the State, to impose and enforce the laws of justice by “protecting, as far as possible, every member of the society from the injustice or oppression of every other member of it” (Smith WN, IV.ix.51).

In a fairly brief but important chapter on the secularization of political economy Halteman and Noell trace some of the ways by which economics was changed from a moral science into more of a natural science during the nineteenth century, using Ricardo, Malthus, Bentham, and Marshall as primary examples. The authors make the insightful observation that methodological discussion in economics is often a discussion of the degree to which intentional human action, i.e., policy, can alter economic circumstances. If the market is structured by nature and not by man, then moral reflection and dialogue is unnecessary and policy is of no use. For example, the authors show how Bentham looked at demand as structured by nature’s placing “mankind under the governance of two sovereign masters, *pain* and *pleasure*”;¹ and how Jevons and the other marginalists showed that value derived from the natural principle of diminishing marginal utility rather than from moral reflection.

The analysis of Marshall’s work shows how, although he was uneasy about the separation of economics from moral reflection and did not view economics as a value-free enterprise; his work became the foundation of the formal, mathematical approach to economics which isolated itself from moral reflection. While the authors correctly point out that Marshall’s work has been used to promote an approach with which he would not completely agree, they do not expand on this topic and the detriment to the discipline that resulted. That the ‘Great Minds Series’ edition of Marshall’s *Principles of economics* completely omits Book VI on “The distribution of national income” in which Marshall discusses most his policy prescriptions, shows to what extent Marshall has been turned into an advocate for positive economic science.

In the following chapter, Halteman and Noell use the examples of Marx, Veblen, and Hayek to show how moral reflection continues to be an important part of the heterodox tradition. The discussion of the relationship between Veblen and Hayek, and their mutual understanding of institutional development as an evolutionary process, is especially

¹ This is the opening sentence of Jeremy Bentham’s *An introduction to the principles of morals and legislation* (1907 [1789]).

insightful. The authors further show how Hayek engages in reflection on the origins of moral norms, but how he, ironically, sometimes refuses to engage explicitly in moral reflection himself, resulting in inconsistencies in Hayek's evaluation of economic theory.

The final three chapters discuss rational expectations theory and show how a broadening of economic discourse can and has brought moral reflection back into the center of economic thought. The authors show how mainstream twentieth-century economists constructed a mathematical system requiring numerous simplifying assumptions and presuppositions, culminating in rational choice economics. They claim that 'rational choice', as the primary and dominant concept of economic theory, has crowded out complementary considerations that would enhance the study of human provisioning. The authors argue for enlarging the economists' tool kit to include concepts from other social and natural sciences, in addition to rational choice theory, in order to consider a wider range of motivations. Adding such concepts as social norms, psychological tendencies, habits, risk-taking, and customs to the tool kit helps to integrate context, understanding, and the possibility of moral reflection into economic thinking. In order to once again include moral issues in economic analysis the authors argue for interdisciplinary research, such as using the resources of neurobiology, psychology, and sociology, to investigate the efficiency of law, cooperation, and the nature of trust in behavior. The work of Paul Zak (2012), such as his discovery of 'the molecule that makes us moral', seems an especially good example of what the authors propose. The concluding chapter sketches a framework for how an interdisciplinary approach to understanding a socio-political economic system can enhance the way we understand economic decisions.

The one important economist who is not much considered in the book is Maynard Keynes, and this is a serious omission in a book dealing with the role of moral reflection in economics. Keynes made a clear distinction of his view of economics from what he called the natural science view advanced by Lionel Robbins in *An essay on the nature and significance of economic science* (1932). In the summer of 1938 Keynes and Harrod corresponded in a series of letters in preparation for Harrod's presidential address to the British Association, given in August, 1938. Keynes wrote "[i]n the second place, as against Robbins, economics is essentially a moral science and not a natural science. That is to say, it employs introspection and judgments of value"

(Keynes 1971-1989, vol. XIV, 296-297). He argued further, in a later letter, that:

I also want to emphasise strongly the point about economics being a moral science. I mentioned before that it deals with introspection and with values. I might have added that it deals with motives, expectations, psychological uncertainties. One has to be constantly on guard against treating the material as constant and homogeneous (Keynes 1971-1989, vol. XIV, 300).

To sum up, Halteman and Noell's book provides a very broad analysis of how thinking about moral behavior has been an integral part of political economy from the time of the ancient Greeks; how ethical considerations were stripped from economic thinking in the nineteenth century; and how modern interdisciplinary approaches are once again opening the door to including moral reflection in economic thinking. This is an excellent book that will be useful to all economists who view economics as a moral science.

REFERENCES

- Bentham, Jeremy. 1907 [1789]. *An introduction to the principles of morals and legislation*. Oxford: Clarendon Press. <http://www.econlib.org/library/Bentham/bnthPML.html> (accessed April 2013).
- Keynes, John Maynard. 1971-1989. *The collected writings of John Maynard Keynes*, vols. I-XXX, ed. D. Moggridge, Royal Economic Society. London: Macmillan.
- Robbins, Lionel. 1932. *An essay on the nature and significance of economic science*. London: Macmillan.
- Skidelsky, Robert. 2009. *Keynes: the return of the master*. New York: Public Affairs.
- Smith, Adam. 1976 [1776]. *An inquiry into the nature and causes of the wealth of nations* [WN]. Oxford: Oxford University Press.
- Zak, Paul. 2012. *The moral molecule: the source of love and prosperity*. New York: Penguin Group.

Betsy Jane Clary is professor of economics at the college of Charleston and a visiting professor at EM Strasbourg, University of Strasbourg. Her current research interests include the relationship between ethics and economics, the history of social economics, and financial market instability. She has most recently published in the *Review of Social Economy* and the *American Journal of Economics and Sociology*. She is a past president of the Association for Social Economics. Contact e-mail: <ClaryJ@cofc.edu>

Review of Ryan Walter's *A critical history of the economy: on the birth of the national and international economies*. London: Routledge, 2011, 138 pp.

TILL DÜPPE

Université du Québec à Montréal

During the 18th century, when Britain was virtually in continuous war, the British fleet got part of its naval stores from the Baltic States. This trade was essential for maintaining Britain's military strength. Consequently, when Denmark joined Britain's enemies during the second Anglo-Dutch War and cut off this trade-route, finding a solution to the trade disruption was key for winning the war. Economic wealth and military strength were thought to be the same thing—thus, according to Ryan Walter, early modern man could not conceive of the economy. But when Adam Smith spoke about wealth and thought of labor, and when Ricardo spoke about wealth and thought of class conflicts, Walter argues, the economy was born. Only with the development of “forms of arguments” about trade that remain silent about the strength of the state, can we think of such a thing as the economy.

Walter develops this claim in a book of only 120 pages. It is based on his doctoral thesis at the Centre for the History of European Discourses at the University of Queensland. The book contributes to the *Routledge/RIPE Studies in Global Political Economy*. According to the editors, the series aims to show “the inseparability of economic from political, social and cultural questions”. If the notion of a separate economy is so terribly mistaken, as the folklore of the critique of political economy goes, the critical question to pose is indeed, how did it come about that people began believing so in the first place? And so Walter achieves the opposite of the series' goal: revealing the arguments that made it possible to separate the economic and the political. Why this happened, and whether it was a good thing, as the intended reader of the book series presumably wants to know, is not Walter's concern. He is instead interested in the *forms of argument* that separate the two domains. He thus presents an epistemic morphology that interrogates the construction of the arguments that make the economy

“cognizable” as a “nominal field” distinct from the state and a domain in and of its own.

Doing so is more than conceptual history, Walter believes, for he stakes out two positions that make his narrative “critical”. He is against the “naturalism” of the economy (the economy is not a natural object that has always been there) and against the “progressivism” of economic science (political economists did not see better, but differently, than other writers before them). Both are obvious requirements for speaking about the birth of the economy. The belief in the existence of a domain of the economy depends on certain “practices of knowing” that play out in certain discursive practices and, ultimately, certain practices of governing. Walter promises to show the reciprocal relation between intellectual arguments and governmental practices. He does not ask whether or not an argument is true, but what it does, what “nominal fields of action” it creates. This allows him to see political economy not as a result of progress, but as emergent from certain “shifts of specific genres of intellectual argument” which remain invisible if one uncritically accepts Whiggish writings of history, such as Smith’s distorted account of mercantilism.

The book keeps only some of these promises. Despite his critical intention, Walter’s reconstruction of early modern economic writings reasserts much of canonical intellectual history with only little insights into discursive practices.¹ But before assessing his strategy, it is worth outlining the contents of the work. The book is very neatly organized. The first part describes the intellectual context of the birth of the economy (in early modern British writings on “counsel on trade”). The other three parts explain how this birth took place: one before the economy was conceived; another on its half-way form (Adam Smith); and a last on its final production (David Ricardo). Each of the last three parts is symmetrically divided between a chapter on the national economy and another on the international economy—a division that is thus not itself historicized. In the first two parts, on pre-Smithean writings, Walter goes through a considerable number of pamphlets one by one. The last two parts are closely based on Smith’s and Ricardo’s canonical texts.

¹ The literature to which Walter refers is different from what this reader expected. He refers to Foucault, Polanyi, McKenzie, Tribe, Pocock, and Skinner, but neither to standard texts on early modern thought, such as by Hutchison and Magnusson, nor to historical studies of early modern science, such as by Shapin, Porter, and Poovey.

The first and most descriptive part of the book pleased this reader most. It describes the context in which new forms of arguments could be developed, and also introduces the two main protagonists: the counselor and the statesman. The beginning of the story is well-placed, upon the background of Italian humanism that had just arrived in Britain bringing with it a new virtue for the British governing class—“being learned”. In order to claim credible interest in the advancement of society one was supposed to be informed. Statesmen and their Privy Council sat together talking over what were the best laws to strengthen the power of the church, of the fleet, and of the merchants. Counselors addressed their advice to statesmen in a respectful, patriarchal fashion. Counsel was neither real debate nor actual command, such that counselors ran into problems with parliamentarians, who also liked to have a word when kings swung their scepters.

When counseling becomes a public affair there are soon others who participate in the discourse—laity, nobles, merchants, and so on—which would make epistemic problems of counsel omnipresent. Recall the work of Steven Shapin (to whom Walter makes no reference), who showed how great the problems of credibility were for the merchant class, as they, for most of Western history, were presumed to make a living by lying (Shapin 1994). Yet Walter did not want to write a book on actual counselors and statesmen since he considers the difference between these two characters only as a “trope” independent of their historical reality. Walter therefore looks closely *into* the texts to see how the trope of the statesman was little by little replaced by arguments independent of administrative agency. In order to do so, Walter distinguishes between different *Genres of Counsel* to guide his reconstruction of the “forms of arguments”. First, the “analysis of interest” aims at understanding the place of the state in a complex geopolitical situation. Second, “political arithmetic” provides means for estimating the strength of the state, and the share that different trades contribute to it. Third, “counsel on trade” is concerned with identifying the sources and potential for growth. The “object” that can be “cognized” by these three genres is the administrative nation state, but not the economy.

In the second part Walter explores more deeply the argumentative link between state strength and wealth in the counsel on trade genre. He convincingly sketches an image of anxiety in a time of endemic warfare: “Our counselor understood the political world in which the

state existed as a threatening terrain of ambitious princes and confessional conflict that produced extended wars and uneasy truces” (p. 50). The arguments that evolved in this situation involved, for example, typologies of trade broken down by their effects on the balance of power (for example, fishery was promoted as productive food source for sailors), explanations of the strength of the Dutch (their control of water, extensive fortifications, supply of ships and munitions, and so forth), arguments about colonies as a means for managing the balance of power, and of course the inglorious equation of economic wealth with bullion to pay the army. At a more analytical level, Walter analyzes problems with the circulation metaphor by which trade was conceived as a whole. This part of the book makes a truly original contribution by showing that counsel on trade was its own genre and not only a stopover between sovereignty and political economy.

Only since Smith, Walter argues in part three, have we learned to speak about wealth in a more abstract fashion. Introducing an “analytical wedge” between wealth and strength, Smith was the midwife of the economy, but did not give birth to it. “Critical history” in this part amounts to no more than showing the strategic element in Smith’s reconstruction of mercantilist thought, which has been long recognized. Smith, Walter argues, “ignored the geo-political character of counsel on trade and construed its argument as failing to analyze the allocation of capital as set down in his system”. Although subverting Smith’s rhetorical break with his predecessors, Walter confirms the relative novelty of his arguments and thus the canonical idea of Smith being the founder of political economy—though not quite the founder of the dismal science. His presentation of Smith goes little beyond textbook presentations: wealth is defined with reference to a homogenous category, *labor*—a definition abstract enough to not immediately lead to thoughts of bullion, munitions, or naval stores. Rather than the statesman, the moral agents of his discourse are individuals acting out of self-love and thereby bringing about an efficient allocation of capital goods. In addition to this static analysis, Smith’s theory of capital and the difference between productive and unproductive labor is given space. The military appears no longer as the manifestation of wealth, but as unproductive labor. National security plays the role of an exception to the rule, rather than the main focus of the analysis of wealth.

Regarding the international economy, Walter argues that Smith indeed conceived of the “nominal object” of the “world economy”—“the whole globe of the earth, of which the wealth, population, and improvement may be either gradually increasing or gradually decaying” (Smith, WN, IV.3.45). Smith thought in terms of one “mass” of world wealth rather than a structure of power enabled by nationally specific sources of wealth. The ‘system of natural liberty’ results in an optimal allocation of capital directed to overseas trade, an idea that resulted in a new yardstick for national trade policy: neutrality. That was inconceivable in previous writings in counsel on trade as it is insensitive to the balance of power between nations. And so Walter writes justly that with Smith the very genre of counsel on trade begins to degenerate, since a neutral trade policy deprives the counselor of the possibility of actually counseling the statesmen. However, despite this “dislocation of argument in counsel on trade” (p. 91), the implied audience of Smith is still the statesman with the power to increase wealth by specific laws. The traditional role of the statesmen is only “partially eclipsed” (p. 76) by the moral agent.

It was Ricardo, in the early 19th century, who, according to Walter, gave birth to the economy. Again, Walter’s presentation differs little from others. He goes through the deduction of the functional distribution of income from the law of diminishing returns in agriculture. The abstract notion of economic “classes” (landlords, capitalists, and workers) was most decisive, according to Walter: “The emergence of class interests founded by distribution represented the most important break with *Wealth of Nations* and with the preceding forms of argument in counsel on trade” (p. 96). The statesman consequently vanishes from Ricardo’s writings altogether and with him the problem of national strength and inter-state rivalry. International trade was a matter of diverging growth paths according to different returns in agriculture.

The humanistic image of counsel offered by a subject to a sovereign was replaced by the imperative to disseminate the scientific insights of political economy to a legislature; Ricardian political economy was not a branch of the statesmen craft, but a science in its own right.

Notably, Ricardo treated the international and the national economy symmetrically, such that there is in fact no analytical difference.

Indeed, one can only speak of a difference between the national and the international economy if the “nation” supersedes the economic. The difference is political. Walter draws the correct conclusion: “political economy had emerged not only as a science of wealth with little to say about power, but as one that could look dubiously at the very existence of the nation state” (p. 111). But this is also why the difference between the national and international economy should itself be historicized rather than taken for granted, as in this book.

Three questions remain open after reading Walter’s book. First, this reader would have liked to learn about the reasons why these forms of arguments emerged. Why was the economy separated from the political? Is it not the task of “contextual intellectual history” to explain why certain arguments emerged at a particular point in time and became, or failed to become, convincing? To be sure, refusing to rule on the truth or falsity of an argument is a first step towards, but still a far cry from, historical criticism. Walter works patiently from text to text but gives relatively little consideration to why these texts succeeded. Ricardo, for example, may have managed to silence much of the statesman trope, but the interesting question is how his argument nevertheless became effective in the political arena. The big topics in historical epistemology, such as credibility and authority, are strikingly absent from Walter’s account. How did intellectual arguments emerge from, and affect, historical life? Walter leaves us with a description of intellectual change in which forms of argument emerge and recede without reasons and with no consequences.

Only once one has answered this question, can one address a second open question: what happened to this separation after its initiation? Was it ever fully accomplished? The separation of state and wealth, Walter notes in passing towards the end, could indeed be exploited by both the left and the right. Mill, Marx, Marshall, down to Arrow and Debreu all had their own ways of dividing the economic from the political. We certainly do not think of the economy today as a separate object because we implicitly trust in Ricardo’s notion of “classes”.

The third open question must be the most pressing for the reader of this book series: what does this birth tell us about the current state of the linkages between the political and the economic? After reading Walter’s account, the contemporary folk-belief that the economy is all about power-relations remains a riddle; and likewise the persistent folk-justification for the existence of economic science: that it is “politically

relevant". In other words, the lessons that can be drawn from Walter's analysis of the forms of arguments that separated the economic from the political remain unclear. In the very last paragraph Walter gives a vague hint. He mentions Afghanistan and Iraq, arguing that the separation between wealth and strength is valuable as they are so closely intermingled in present-day discourse. "The politics economics split provides a valuable ideological function in contemporary political life" (p. 118). It is that "valuable function" about which this reader would have liked to learn more from a book on the birth of the economy.

REFERENCES

- Shapin, Steven. 1994. *A social history of truth: civility and science in seventeenth century England*. Chicago: University of Chicago Press.
- Smith, Adam. 1976 [1776]. *An inquiry into the nature and causes of the wealth of nations* [WN]. In *The Glasgow edition of the works and correspondence of Adam Smith*, Vol. 2, eds. R. H. Campbell, and A. S. Skinner. Oxford: Oxford University Press.

Till D ppe is an assistant professor for the history of economics at the Universit  du Qu bec   Montr al.

Contact e-mail: <duppe.till@uqam.ca>

**Review of Samir Okasha and Ken Binmore's *Evolution and rationality: decisions, co-operation, and strategic behaviour*.
Cambridge: Cambridge University Press, 2012, 281 pp.**

WILJAN VAN DEN BERGE
Tilburg University

There is a long history of sharing models between evolutionary biology and economics. Most famous is of course the introduction of game theory from economics into biology, which was then transformed into evolutionary game theory and re-imported into economics (Grüne-Yanoff 2011). Another area where both biology and economics have things to say is on cooperation. Both evolutionary biology and rational choice theory (RCT, hereafter) have problems accounting for cooperation within their standard frameworks. This has led to the development of a literature on cooperation that straddles the borders between economics and biology, and also philosophy.

These topics were explored in two conferences organized by Samir Okasha and Ken Binmore at the University of Bristol, and eleven papers from these conferences are collected in this volume. In the introduction, Okasha and Binmore emphasize the strong connections between evolutionary theory and RCT. Continuing a theme from Okasha's recent work (Okasha 2007; and 2011), they focus particularly on how both RCT and evolutionary theory conceptualize behaviour 'as if' it is maximizing some quantity: utility in the case of RCT and fitness in the case of evolutionary theory. They suggest that the main question arising from this is, "when is it possible to identify the economist's notion of utility with the biologist's notion of fitness?" (p. 2).

While this specific question might be of interest to some, it makes the book much more worthwhile that not all the chapters explicitly address it. Rather, the authors discuss a range of conceptual and methodological issues regarding the many interrelations between evolutionary theory and RCT, construed in their broadest senses. The most prominent interrelation is the use of game theory in both economics and biology to model strategic interaction. In addition, about half of the contributions address issues of cooperation in one way or another. Different explanations of cooperation, such as strong

reciprocity (Vromen, Gintis and Sterelny), team reasoning (Gold), or the evolution of preferences (Berninghaus, et al., and Wolpert and Jamison) are discussed. In this review I will try to give a flavour of the wide range of contributions in the book. I will divide them between the themes of *evolution and rationality* (broadly conceived), and *cooperation*.

I will begin with the two chapters on evolutionary game theory and then discuss some other ways of thinking about the relation between evolution and rationality. The first chapter, by Peter Hammerstein, provides a short introduction to game theory as it is used in biology that would be useful to those who are unfamiliar with the major developments in game theory. He discusses the main solution concepts and their potential usefulness in the social and biological sciences. Hammerstein argues that we should interpret players that play an equilibrium strategy (i.e., each choosing a strategy such that no player has an incentive to deviate from their chosen strategy) as resulting from learning that this strategy is most successful, rather than from strong rationality assumptions. He expresses his hope for a more explicitly Darwinian theory of learning.

In their chapter, Simon Huttegger and Kevin Zollman discuss the limits of what they call “ESS methodology”. An evolutionarily stable strategy (ESS) is one which, once generally adopted by a population, cannot be successfully invaded by any alternative strategy. Huttegger and Zollman criticise the standard method, within evolutionary biology, of only finding out whether a given strategy would be an ESS, but not considering the evolutionary dynamics that underlie the equilibrium. This can lead to mistaken conclusions. For example, it could mean that we only consider an equilibrium that is very fragile, or that is highly unlikely to be reached. They argue for considering the dynamics of the system, as epistemic game theory does in economics.

Another question that often returns is whether our evolved behaviour is in some sense ‘optimal’. In their chapter, Claire El Mouden, Maxwell Burton-Chellew, Andy Gardner, and Stuart West remind us that the biologist’s assumption of fitness maximizing behaviour is really only that, an assumption. In practice, we should at best expect to find adaptive behaviour. They discuss many reasons for this, such as developmental restrictions on organisms and evolutionary forces other than natural selection. The same holds true for the assumption of optimal behaviour by humans. They argue that many of the reasons why we do not find optimization in evolutionary biology also hold in

economics. In his chapter, Alasdair Houston looks at one particular violation of optimal ('rational') behaviour, the violation of transitivity. He argues that it can be given an evolutionary explanation, for example in terms of the evolution of behavioural rules.

Henry Brighton and Gerd Gigerenzer look at another aspect of rational behaviour: are we rational in "large worlds" as well as "small worlds"? Small world problems are ones whose characteristics can be perfectly known, such as a lottery. In contrast, large world problems require coping with inherent ignorance, which turns uncertainty into risk. Leonard Savage, who coined the distinction, famously said that it was "utterly ridiculous" to use his theory to try to understand large world problems (Savage 1954, 16). Brighton and Gigerenzer focus on inductive inference. They distinguish between the internal model that an agent holds and the external data-generating process. An agent's internal model can diverge from the real world on which it is supposed to be based due to stochasticity, underspecification, and misspecification. As decision problems acquire more 'large world' characteristics, these sources of error play a larger role and it becomes difficult to find the optimal strategy. We can, however, look at what simple rules have worked in the past and gradually improve those. This is the "relativist strategy" that Brighton and Gigerenzer advocate.

Siegfried Berninghaus, Werner Güth, and Hartmut Kliemt discuss a different way of understanding the relation between evolution and rational behaviour. They argue that we should distinguish between long and short-run time-scales. The long term time-scale is where preferences evolve (which they characterise as "push", since it does not involve goal-directed agency); the short-run scale is where we find rational behaviour ("pull"). Thus, the objective payoffs that determine evolutionary success are mapped onto the subjective payoffs that determine behaviour in the 'games' we play in everyday life. At the proximate level, we can have behaviour that is for example motivated by a social norm. But in the long run this will tend to produce higher objective payoffs for the individual.

I will now turn to the contributions on *cooperation*, starting with one that takes issue with Berninghaus, and co-authors' view. David Wolpert and Julian Jamison argue that indirect evolutionary models of the sort Berninghaus, et al. propose are too complex and can therefore only handle very simple situations. We should rather understand the long-term process as a learning process and the short-run process

as choosing a subjective utility function. This choice is analysed as adopting a “persona”, that we commit to and that will determine our actions in a given situation. Over time we learn which personas lead to the best objective results in given situations. This can explain short-run cooperation, where people adopt a cooperative persona in a particular context, by pointing to the long-run consequences of this behaviour.

Natalie Gold discusses another explanation for cooperative behaviour: team reasoning. The basic idea is that instead of asking “what should I do?” players will ask “what should we do?” and then play their part in the team strategy. How team reasoning exactly works depends on the assumptions we make about the payoffs for the team and how people identify with a team. Views diverge on this. Robert Sugden argues that people need explicit agreements, whereas Michael Bacharach argues that cues from the environment induce a so-called “we-frame” of thinking.

Gold examines the difference between the team reasoning account and the idea of other-regarding preferences. She finds that both have similar mechanisms—objective payoffs diverge from the subjective perception of the game—but she argues that team reasoning involves a more fundamental transformation of agency. She concludes with an unconvincing discussion of which explanation is better. One criterion she proposes is that we have a plausible evolutionary story for the team reasoning account. But it would be very easy to develop such a story for the other-regarding preferences account and then we are still left with deciding which story is better. It seems to me that carefully designed experiments are a more promising approach, since there we can at least partly control for the characteristics that induce team reasoning (framing or explicit agreements) and other-regarding preferences.

Another recently proposed explanation of cooperation is that humans are strong reciprocators. Jack Vromen tackles this hypothesis. In particular he addresses the confusions surrounding the concepts of evolutionary and psychological altruism and selfishness. For example, strong reciprocity theorists argue that ‘strong reciprocity’ (conditionally cooperating and punishing those who do not cooperate) is genuinely altruistic at both the evolutionary and psychological level, in contrast to ‘reciprocal altruism’, which is selfish on both levels. Vromen clears up these confusions by carefully distinguishing the different notions of altruism involved. He argues that evolutionary selfishness (lifetime fitness maximization) need not imply psychological selfishness.

But perhaps, Vromen argues, it does not really matter what causes strong reciprocity in terms of ultimate desires. What matters are the conditions under which it arises.

Herbert Gintis continues his argument for a synthesis of the behavioural sciences (Gintis 2009), resting on the pillars of “gene-culture coevolution, the sociopsychological theory of norms, game theory, the rational actor model, and complexity theory” (p. 213). Gintis tries to relate a lot of theories and concepts to each other to come up with a synthesis, because he believes that inconsistencies between the social sciences are a huge problem for progress. While one might wonder about a lot of the links that Gintis tries to establish, I am also doubtful of the general usefulness of this exercise. It does not necessarily strike me as a problem that, for example, sociology offers predictions that are inconsistent with those of economics.

Kim Sterelny provides an interesting take on the same sorts of issues, but from a pluralistic perspective. He argues that it is not a problem that models are inconsistent, since they each highlight different aspects of a given target system. Sterelny compares the rational actor with self-regarding preferences (‘homo economicus’) model with the strong reciprocity model and human behavioural ecology model. He argues that strong reciprocity models are a useful extension of the standard homo economicus model that retains its generality or tractability. However, he argues that human behavioural ecology, which explains human behaviour in terms of a combination of fitness maximizing population-level processes and individual adaptivity, is less applicable to modern societies, since the complexity of modern societies has led to a decoupling of reproductive fitness from “what we want”, our utility.

In conclusion, philosophers of economics might feel that this book focuses too much on evolutionary theory, but I believe that it tackles a lot of questions that are of interest for them as well. It also shows that there is a lot we do not yet understand. For example, some people argue that human behaviour can be understood as optimal in some sense, be it fitness-maximizing or ‘rational’, whereas others doubt this very much. The interpretation and explanation of economic experiments also remains a significant concern, both in terms of violations of the standard rationality axioms and the appearance of cooperative behaviour. How should we explain this behaviour? *Evolution and*

rationality provides strong evidence that biologists, economists and philosophers have a lot to gain from discussing these issues together.

REFERENCES

- Gintis, Herbert. 2009. *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton: Princeton University Press.
- Grüne-Yanoff, Till. 2011. Models as products of interdisciplinary exchange: evidence from evolutionary game theory. *Studies in History and Philosophy of Science Part A*, 42 (2): 386-397.
- Okasha, Samir. 2007. Rational choice, risk aversion, and evolution. *The Journal of Philosophy*, 104 (5): 217-235.
- Okasha, Samir. 2011. Optimal choice in the face of risk: decision theory meets evolution. *Philosophy of Science*, 78 (1): 83-104.
- Savage, Leonard J. 1954. *The foundations of statistics*. New York: Wiley.

Wiljan van den Berge graduated from the research master in philosophy and economics at the Erasmus Institute for Philosophy and Economics (EIPE) in September 2012 with a thesis on the use of evolutionary models in the social sciences. He is currently pursuing a Master's degree in economics at Tilburg University.

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

**Review of Axel Gosseries and Yannick Vanderborght's
Arguing about justice: essays for Philippe Van Parijs.
Presses Universitaires de Louvain, 2011, 422 pp.**

JULIANA BIDADANURE
University of York

What better way to pay tribute to a unique and innovative philosopher than by asking his dissenters, disciples, and colleagues to write about his contributions? This is the idea that Yannick Vanderborght and Axel Gosseries—Philippe Van Parijs's colleagues at the Hoover Chair in Economic and Social Ethics at the Université Catholique de Louvain—have carried out for Van Parijs's sixtieth birthday. As they explain in their foreword, "We thought Philippe would prefer a book to a cake. We also wanted to make sure he did not [already] have the book—not an easy task" (p. 35). Reading *Arguing about justice*, one wonders how Van Parijs must have felt reading a book that mentions his name two hundred times. His expression must have fluctuated a great deal—intermittently blushing, smiling, frowning and sighing. The fifty contributors all talk about Van Parijs with admiration, even when their praise paves the way to criticism. But the simple fact that so many writers were willing to participate in the volume and engage with Van Parijs's key ideas tells us much about both his contribution to the field and the personal esteem in which he is held. As a tribute the book is a self-evident success. It does justice to Van Parijs's eclectic oeuvre and testifies to the feelings of friendship he inspires in people.

The edited collection opens with a very helpful foreword by the editors. The forty-one short articles that follow are sorted alphabetically rather than thematically. As a result, locating papers on a single topic—say, universal basic income—requires leafing through papers on linguistic and territorial justice, education and health policy, the media, Marx, gender justice, family policy, the relevant time-unit for equality and freedom, and many more. The collection takes the reader from a discussion of the challenges the internet raises for journalism (Bruce Ackerman), to a paper on Van Parijs's analogy between jobs and marriages (Anne Alstott), to a reflection on the implications of the whole life view for designing distribution (Ian Carter), to a paper exploring

our tendency to blame and stigmatise survivors (Jean-Michel Chaumont). Many articles are particularly remarkable for their original approaches to their objects of study—such as the portrayal of a society in which Marx and Freud never lived (Jon Elster), a simulation involving turtles being allocated an unconditional basic sugar income (Paul-Marie Boulanger), and a discussion of cities as worthy objects of political philosophy (Daniel Weinstock).

Through this broad diversity of topics and methods, the articles paint a multi-faceted and rich picture of the eclectic philosopher to whom their essays are dedicated. Overall, the articles are refreshing, original, thought-provoking, and sometimes confounding to those who endorse Van Parijs's approach. The shortness of the contributions has an interesting incentivizing effect: one is more tempted to read articles one would otherwise ignore. This incentive is further reinforced by the fact that the abstracts printed above each article are in many different languages (including French, Dutch, Catalan, and Esperanto). This nod to Van Parijs's recent interest for linguistic justice also has the side effect of teasing the reader into examining the articles themselves in order to find out what they are about. This intrigue-inducing effect may lead readers to consider topics otherwise dismissed as uninteresting or irrelevant to their own intellectual pursuits.

Writing a review on such mosaic of papers is not an easy task. The main problem is to decide which approach to adopt—zooming in or out. While adopting a general standpoint risks overlooking precious nuggets contained in the articles, zooming in might lead the reviewer to misrepresent the big picture of the collection—including its general coherence and unifying features. For this reason, I have decided on a compromise. In the remainder of this review I will focus on the topic of universal basic income (UBI). I will mainly focus on four chapters epitomizing what I take to be the core theoretical and practical difficulties faced by the basic income movement: the limits of instrumental justifications (Christian Arnsperger and Warren A. Johnson); the tension between ideal and pragmatic approaches (Denis Clerc); the practicalities of implementation (Eduardo Matarazzo Suplicy); and the challenge of mobilising communities and building support (Bill Jordan).

Arnsperger and Johnson offer a defence of UBI “as an equal opportunity tool in the transition toward sustainability” (p. 61). They contend, ambitiously, that *true* equality of opportunity includes

the freedom to produce and live outside of capitalist markets. Since such equality of opportunity is a fundamental human right, we should feel discriminated against when it is denied to us. Endorsing a Marxist conception of human flourishing that regards our productive and creative nature as essential, they call for a basic income to enable non-capitalist experiments:

the real freedom—to freely choose not just some intra-capitalist life style (e.g., becoming a marketing agent rather than a bank director, or creating one's own capitalist software company instead of working for Google) but to choose between an intra-capitalist way of life and an extra-capitalist one (e.g., moving to an ecovillage and exchanging goods and services within a network of user[s] of mutual-credit currency, instead of staying in the hyper-competitive agrochemicals company with whose salary one can consume all one's fill) (p. 63).

Arnsperger and Johnson argue that 'marginal' people who explore frugal modes of living are currently unfairly stigmatised. On the contrary, they should be encouraged as pioneers since they experiment with the humanly and environmentally sustainable lifestyles that societies need to transition to. The authors' version of the UBI—the Economic Transition Income (ETI)—is meant to promote such alternative experiments.

While the first line of Arnsperger and Johnson's argument is a radical but fair interpretation of the requirements of Van Parijs's real libertarianism, the second line of their argument seems to depart from it. On real libertarianism, UBI should be truly unconditional and not aimed at supporting certain kinds of behaviours that make us all better off. The authors' view that "the will to experiment needs to be triggered" (p. 68) thus seems to conflict with the very extensive notion of freedom they defend, and also, surely, with Van Parijs's conception of 'real freedom' as the freedom to do whatever one might want to do.

One could argue that this is merely a theoretical problem: in the political struggle to implement UBI, a variety of non-necessarily consistent arguments might point in the same direction. The persistent problem, however, is that the two theoretical underpinnings can lead to very different forms of basic income being put into practice. While we may welcome the implementation of a basic income for those who experiment with sustainable economic alternatives, we lose out a great deal if *only* such activities are considered as deserving a UBI. In other

words, if Arnsperger and Johnson really believe in UBI, they must be careful that their proposal provides *additional* reasons to support UBI instead of threatening its foundations.

Clerc shares an interest in the ideal of UBI with Arnsperger and Johnson, but he expresses strong doubts about the capacity of such a “big idea” to change society. Such radical and utopian projects, he argues, risk exerting little to no impact on societies because they seem inaccessible and because no politicians will want to take such risks. He explains:

In the face of resistance and criticism, Philippe Van Parijs (and Yannick Vanderborght) have bravely defended BI, but they did so by increasing the depth and refinement of their project, rather than by making it more accessible in debate with their critics, and thus condemned themselves forever to sow on stony ground (p. 169).

Clerc raises interesting and fundamental questions about the role of ideal principles. He argues that supporters of UBI should only fight to bring about a less unjust world. To achieve this modest goal one should support incremental reforms that pave the way to a better tomorrow, rather than big ideas. To illustrate his point, Clerc presents the French welfare reform *Revenu de Solidarité Active* (RSA) as an improvement over the former guaranteed minimum income, and as an example of the type of reforms defenders of UBI should fight for. He concedes that RSA is far from the UBI ideal, but argues that “we are getting closer” (p. 171).

I disagree with Clerc’s analysis of RSA as an improvement in the direction of UBI, and I think his own example shows why we do need big ideas. The former French guaranteed income support scheme (RMI) was a means-tested allowance of less than 500 Euros provided to persons without other sources of income. The RSA, on the other hand, comprises two allocations. The first is similar to RMI, but includes sanctions on those who do not fulfil their obligations or reject two ‘suitable’ job offers. The second is an allowance that tops up the income of those in work to help them reach a more decent monthly income. As a result, there is a clear incentive to work since recipients of RSA are always better off when they work, which was not the case with RMI.

Perhaps Clerc means to suggest that RSA is an improvement because it supports poor workers and therefore breaks the traditional divide between in work welfare-producers and out of work welfare-recipients. However, on many other criteria, RSA does not seem to be an

improvement. The new emphasis on sanctions for those who refuse to accept jobs is a backwards step on the path to UBI. Not only does RSA emphasize the duty of citizens to take on a job, even if it is precarious and demeaning, it also perpetuates the view that society should only support those who deserve it. Moreover, desert is defined solely in terms of paid employment. RSA is also based on the view that unemployment is a problem of individual commitment, even when society simply does not provide enough employment opportunities. Lastly, it accommodates rather than condemns the unacceptable fact that some full-time jobs are so inadequately paid that they do not allow one to make a living. As a result, vulnerable individuals are induced to accept demeaning and exploitative positions. As insufficient and unsatisfactory as the RMI was, it seems that this reform has not moved France closer to the real freedom of UBI.

For these reasons, RSA does not seem a suitable example for the point that Clerc is trying to make, and this also suggests problems in his conception of what UBI's proponents should fight for. Van Parijs and Vanderborght do welcome advances towards UBI, but they do so while still promoting the ideal of a *truly* universal basic income. It is only by reference to the ideal that one can identify whether a given reform really is an improvement. Does the income support reform empower citizens—as Karl Widerquist (p. 387) would rightly ask—by making them freer to say no to oppressive, coercive, and undignified forms of employment? Will it enable them to flourish, without stigmatization or resentment? UBI is a radical reform that could potentially transform the lives of the most stigmatized and vulnerable. Only reforms that really aim at bringing us closer to this goal deserve to be fought for.

Suplicy's paper also touches upon the fundamental challenges of implementing UBI. The author paints an optimistic picture of how basic income can serve as an instrument of development in his analysis of the progress achieved by Brazil's *Bolsa Família* program: "The proportion of poor families [...] decreased from 39.4% in 2003 to 25.3%, in 2008 [...] The 20% poorest families had an income per capita increase 47% faster than the income of the richest 20%" (p. 339).

Suplicy shows convincingly that the modest *Bolsa Família* program has significantly reduced poverty. He also demonstrates that the country is engaged in a real process to bring about a citizen's basic income. Suplicy acknowledges that Brazil is still one of the most unequal countries in the world, but shows that the UBI project is slowly

contributing to making the country less unjust.

Jordan's paper addresses the key challenge of building political support for basic income. Written from a British perspective, his article summons up the history of the campaign for a basic income in the 1970s in a small industrial town in Southwest England. He describes the mixture of very different groups fighting for UBI at the time: workers who had been made redundant, Trotskyist trade unionists, the long-term unemployed, single parents, and disabled persons. The main strength of Jordan's paper is that it persuasively demonstrates that advocates of basic income should work to build support for the policy among the most disadvantaged. He appeals to Guy Standing's concept of the 'precariat' to explain why and how UBI defenders should turn to the excluded. He makes a very interesting point by tying the struggle for a UBI to the difficulties faced by young people, and argues for entrusting them with this fight.

The elderly, the unemployed, the former activists, the carers, the 'marginals', and the environmentalists—as Arnsperger and Johnson suggested—should all rally behind the struggle for a universal basic income. The various papers I have discussed all suggest that UBI would benefit a broad variety of individuals and that these groups should be approached more actively. Yet many will be opposed to UBI, and this is precisely where those who argue about justice have a role to play: by providing the arguments and the empirical evidence to support the cause. This is the answer that many authors seem to offer to John Baker's question "If you're an egalitarian, why are you devoting your life to arguing?" (p. 81).

It is fundamentally important to understand why so many people who would benefit from UBI still oppose the idea, and to win them over. The ideological opposition to UBI—such as the dogmas of individual responsibility and market-wages as moral desert (Bidadanure 2012)—needs to be theoretically undermined. In general, the articles in this collection offer a very strong interpretation of what *arguing about justice* means: contributing to the theoretical and practical struggle for change. As John Roemer, for example, puts it, "the most important problem for the social science of inequality is understanding how electorates have come to *acquiesce* to policies which increase inequality, and to try and reverse this acquiescence" (p. 301).

Universal basic income, I would argue, is a truly radical and yet realistic idea. It appeals to the power of utopia and the strength of

pragmatism, and is thus at once inspiring and convincing. UBI is certainly a compromise, but does it surrender to capitalism, as the philosopher Slavoj Žižek claims (2009)? Žižek argues, against Van Parijs, that, as nice as it sounds, UBI is not a solution because it only sweetens capitalism instead of threatening its foundations. Žižek illustrates his view by quoting Oscar Wilde who, in *The soul of man under socialism* (1891), composes the following portrait of charity:

[T]his is not a solution: it is an aggravation of the difficulty. The proper aim is to try and reconstruct society on such a basis that poverty will be impossible. And the altruistic virtues have really prevented the carrying out of this aim. Just as the worst slave-owners were those who were kind to their slaves, and so prevented the horror of the system being realised by those who suffered from it, and understood by those who contemplated it (Wilde 1891; quoted in Žižek 2009).

I think that Žižek is mistaken. Endorsing UBI does not amount to surrendering to capitalism. Far from maintaining people in situations of poverty and alienation, a universal basic income empowers individuals: freeing them from dependence on those who own the means of production, and bringing opportunities for emancipation and self-determination.

Can we then go so far as considering UBI “the capitalist road to communism”, as Van Parijs once did? UBI can surely take us to a more democratic and equal society. It is far more ambitious to claim that UBI alone will take us to a non-capitalist future, let alone to a communist society. However, anti-capitalists should not see this as a weakness of UBI. The struggle for basic income is fundamentally a struggle over and for time: the time to refuse exploitative options and seek alternative occupations; the time to flourish and care for others; the time to be outraged by injustices and participate in the political struggles of our times. As a result, it is very likely that a truly universal basic income can take us a long way in the struggle against capitalism. In fact, if one understands capitalism as both creating and relying upon a state of mind of perpetual insecurity in the vast majority of the world’s population, then it seems that UBI could strongly contribute to overthrowing first, the state of mind on which capitalism depends, and then, by a domino effect, the entire econo-political edifice. If communism is likewise understood in terms of a state of mind of

freedom and emancipation, then perhaps, rather than the capitalist road to communism, UBI may be seen as a communist road out of capitalism.

REFERENCES

- Bidadanure, Juliana. 2012. Short-sightedness in youth welfare provision: the case of RSA in France. *Intergenerational Justice Review*, 1: 22-28.
- Guy Standing. 2011. *The precariat: the new dangerous class*. London: Bloomsbury.
- Slavoj Žižek. 2009. First as tragedy, then as farce: the economic crisis and the end of global capitalism. RSA (Royal Society for the encouragement of Arts, Manufactures and Commerce), 24 November 2009, London. <http://vimeo.com/8073858>

Juliana Bidadanure is a PhD student in the School of Politics, Economics, and Philosophy at the University of York (UK). In 2012, she visited the Hoover Chair at Université Catholique de Louvain for three months. She writes within the field of political philosophy and is particularly interested in questions of social justice in general, and in theories of equality in particular. She currently focuses on questions of justice between age groups, including what it means for people of different ages to stand as equals and the idea of a universal basic income for young people.

Contact e-mail: <jub500@york.ac.uk>

PHD THESIS SUMMARY:

**Evolutionary foundations of Coasean economics:
transforming new institutional economics
into evolutionary economics.**

MASAHIRO MIKAMI

PhD in economics, March 2012

Hokkaido University

In recent years, interest in the role of institutions in economic society has increased and economists have been forced to incorporate them into economic analysis. The institutional perspective has not only brought new challenges for economics, but also led to a re-evaluation, redefinition, and transformation of the discipline itself. It would not be an overstatement to say that the method by which we deal with and formalize institutions determines the boundaries among the various schools of economic thought, and between economics and related disciplines.

New institutional economics (hereafter NIE), is one of the fields that have fluctuated between such boundaries. The work of the official founder of NIE, Ronald Coase, is an appropriate starting point for considering the characteristics of the field and its prospective development. This historical and methodological study explores the development of Coase's economic thought from the 1930s through the 1970s and into recent times. I show that Coase's theoretical orientation transformed around the 1970s from a neoclassical to an anti-neoclassical stance; a fact that points to the evolutionary foundations of Coasean economics.

Chapter 1 begins with the general relationship among neoclassical, new institutional, and Coasean economics. Invoking Imre Lakatos's methodology of scientific research programmes, NIE is distinguished from neoclassical economics: though they share a theoretical core, the former is positioned as an expansion of neoclassical economics. Then Coase's two facets in relation to NIE are provisionally confirmed: the one as a founder of NIE and the other as a critic of mainstream economics. That is to say, whereas Coase's arguments in "The nature of

the firm” (1937) and “The problem of social cost” (1960) had some elements that could be used to defend and expand the neoclassical approach, after the 1970s he dissented from mainstream economics. I therefore call the latter “Coasean economics”, as distinguished from a mere expansion of neoclassical economics.

Chapter 2 examines the historical development of Coase’s institutional economics in a phased manner. At the London School of Economics (LSE) during the 1930s, under the influence of his supervisor, Arnold Plant, and of LSE’s opportunity cost theory, Coase wrote two interrelated articles: “The nature of the firm” (1937) and “Business organization and the accountant” (1938). Examining their interrelationship confirms that his theories of the firm and of opportunity cost were inseparably related. That had implications later in the 1970s, when Coase’s view of economic agents and institutions diverged from the neoclassical theoretical core and he objected to the concept of rational maximizing agents. Because of his realist methodology (Coase 1977; and 1988a), Coase interpreted institutional costs (transaction or organizing costs) as real costs confronting relevant agents, and his main concern was how agents’ recognition of such costs leads to institutional structuring. Hence, Coase views economic institutions in terms of interrelated institutional costs recognized by economic agents through the price mechanism and accounting systems (Coase 1988b; 1990; and 1992).

Chapter 3 contrasts Oliver Williamson’s view of economic agents and methodological attitudes with that of Coase. I begin with Williamson’s criticism of Coase’s theory of the firm, from which the features of Williamson’s transaction cost economics arose. Examining Williamson’s calculative view of economic agents based on opportunism and bounded rationality, I demonstrate that his strategy is operationalist in seeking falsifiable hypotheses and instrumentalist in emphasizing predictions based on unrealistic assumptions. Williamson regarded transaction costs as the economic counterpart of friction in physical systems and used them merely as an analytical instrument for economists to identify where problems (frictions) reside. (He later shifted his analytical focus to agents’ opportunistic behaviour in relation to the attributes of a transaction.) The source of such instrumentalism (and of neoclassical evolutionary arguments) can be traced back through Milton Friedman’s methodology to Armen Alchian’s evolutionary theory (Alchian 1950), where we can find an “outward-looking” type of evolutionary argument

that concentrates on order formation outside agents by economic natural selection.

The final chapter first reviews Adam Smith's "invisible hand" and Friedrich Hayek's "spontaneous order" arguments as typical examples of evolutionary theories in the social sciences. In contrast, on the surface, Coase's institutional economics seems to be a design theory rather than an evolutionary theory in that he regards institutions as the intentional constructs of economic agents. Indeed, in common with Edith Penrose (1952), Coase warned against the use of biological analogies in economics. Nevertheless, I demonstrate that, although Coase never advocated "evolutionary economics" nor is he regarded as an "evolutionary economist", his institutional economics has two evolutionary features. One arises from his interest in the human nature of economic agents—including rule learning and expectations—as the product of evolution (Coase 1978), and although he did not follow through on this interest, it can lead to a long-term analysis of the relationship between human evolution and institutions. The second feature appears in his theory of alternative institutional structures in relation to costs. Although not in a mathematically formalized manner, Coase actually aimed for a dynamic analysis of interrelated changes in institutional structures (Coase 1998).

Both features can be regarded as an "inward-looking" type of evolutionary argument that delves into the internal structure of agents and institutions, in contrast to Alchian's "outward-looking" argument based on economic natural selection. To fulfil the evolutionary framework of Coasean economics, both these features are required. Its implication is that, in the real world of social institutions, economic policy planning must invoke not a one-time, ready-made design, but trial-and-error processes without any fixed goals.

REFERENCES

- Alchian, Armen A. 1950. Uncertainty, evolution, and economic theory. *Journal of Political Economy*, 58 (3): 211-221.
- Coase, Ronald H. 1937. The nature of the firm. *Economica*, 4 (16): 386-405.
- Coase, Ronald H. 1938. Business organization and the accountant, I-XII. *The Accountant*, October 1-December 17.
- Coase, Ronald H. 1960. The problem of social cost. *Journal of Law and Economics*, 3: 1-44.

- Coase, Ronald H. 1977. Economics and contiguous disciplines. In *The organization and retrieval of economic knowledge*, ed. Mark Perlman. Boulder: Westview Press, 481-495.
- Coase, Ronald H. 1978. Economics and biology: a comment. *American Economic Review*, 68 (2): 244-245.
- Coase, Ronald H. 1988a. How should economists choose? In *Ideas, their origins, and their consequences: lectures to commemorate the life and work of G. Warren Nutter*, Thomas Jefferson Center Foundation. Washington D.C.: American Enterprise Institute for Public Policy Research, 57-79.
- Coase, Ronald H. 1988b. *The firm, the market, and the law*. Chicago: University of Chicago Press.
- Coase, Ronald H. 1990. Accounting and the theory of the firm. *Journal of Accounting and Economics*, 12: 3-13.
- Coase, Ronald H. 1992. The institutional structure of production. *American Economic Review*, 82 (4): 713-719.
- Coase, Ronald. H. 1998. The new institutional economics. *American Economic Review*, 88 (2): 72-74.
- Penrose, Edith T. 1952. Biological analogies in the theory of the firm. *American Economic Review*, 42 (5): 804-819.

Masahiro Mikami obtained his PhD in economics in March 2012 from Hokkaido University (Japan), under the supervision of Professor Makoto Nishibe. The language of the thesis was English. The dissertation committee included Professor Makoto Nishibe (Hokkaido University) as the chief examiner, Professor Shiro Nagao (Meiji University), and Professor Tsutomu Hashimoto (Hokkaido University). Masahiro Mikami is currently senior assistant professor at the School of Business Administration, Meiji University, Japan.

Contact e-mail: <mmikami@meiji.ac.jp>

Website: <www.masahiro-mikami.info>

PHD THESIS SUMMARY:

Beyond green and mainstream: on the normative foundations of environmental policy.

STIJN NEUTELEERS

PhD in philosophy, May 2012

Katholieke Universiteit (KU) Leuven

How should we deal politically with environmental public goods? This dissertation examines three influential political-economic approaches to this question, in order to reveal their commonalities, relevance, and limits. I begin by analysing how environmental problems are understood from a political-economic perspective. Environmental problems are collective action problems over common pool resources. This collective action structure is reiterated on different levels, namely a long-distance spatial, an intertemporal, and an intrapersonal one (e.g., procrastination). The combination of these structural problems makes solutions for (some) environmental problems extremely difficult: it makes for a 'perfect storm' (see, Stephen Gardiner 2011).

In order to handle the wide diversity of approaches to dealing with environmental problems, I started with the most obvious approach, namely the one used by those specialised in environmental issues, environmental scientists and activists. In general, their political approach is based on the natural sciences: they look at the objective features of environmental public goods. As a political theory, this approach largely resembles utilitarianism. Therefore it is no surprise that they are confronted with similar normative and political problems with respect to distributive justice (since distributive issues are invisible in aggregates) and democracy (since there is only one policy priority).

The most significant political claim made by the environmental-scientific approach is of physical limits-to-growth. However, the claim of limits-to-growth, while intuitively self-evident, is hard to make solely on theoretical grounds. Moreover, even empirical limits-to-growth accounts do not always offer much policy guidance. On the one hand, they are too crude because of their aggregative nature and their assumption of continuity (and thus substitutability). On the other, they vary widely in

their conclusions because different indicators deal with different aspects. Therefore environmental policy requires, first, indicators beyond aggregation and continuity. In particular, indicators need to reveal which problems are 'preconditional', such as threatening basic needs or the sheer possibility of cooperation. Second, in order to move from descriptive to normative claims, indicators need also a political theory in order to fill in the idea of 'preconditional'.

In general, an environmental-scientific approach has many difficulties with the idea of plural valuations. The second—economic—approach offers a solution for this. It does not focus so much on the objective nature of environmental public goods, but on their subjective valuation as expressed in individual preferences. Ideally, a method of evaluation offers a way of converting these preferences to a common scale in order to allow for comparisons and decisions. In order to be successful, comparisons need to be made between goods, across persons and through time. This, again, brings forward problems associated with utilitarianism. An account concerned with choosing the policy that maximizes welfare requires converting all information into commensurable units, which need to be homogeneous, anonymous, and non-temporal. These requirements come up against problems of comparability (valuations that are different in kind), interpersonal comparison, and discounting.

However, economists do not use such a pure method of evaluation, but one that uses market prices, namely cost-benefit analysis (CBA), a method which is omnipresent in environmental policy. While CBA works for market goods, the underlying utilitarian problems return if one has to create prices for non-market goods. CBA becomes more controversial the more morally significant differences between separate goods, persons, and moments become. Unfortunately, such differences are often characteristic of environmental public goods. Incorporating such differences seems to demand a more deliberative method of decision-making rather than an aggregative one.

Economic preference-based approaches have problems considering value pluralism (valuations different in kind). The third—political-theoretical—approach focuses on the plurality of value systems or comprehensive doctrines rather than on individual preferences. A political-theoretical approach, such as liberal egalitarianism, is conceived as an answer to the political challenges of distributive justice in a context of value pluralism. It distinguishes between a just

procedural framework and the democratic debate between comprehensive doctrines within this framework, a distinction not made by the two previous approaches.

Here the question is whether particular environmental public goods relate either to the just framework or to the debate between different conceptions of the good. This distinction allows the development of accounts of how to deal with the distributive impact of environmental public goods and of our duties towards the environment. CBA does not make this distinction and considers all problems as a matter of the good, which is a general problem with utilitarianism. However, while a political-liberal analysis can draw the line between the just and the good, thereby providing a framework for decision-making, it is limited with regard to the content and institutions of environmental policy (which are respectively dealt with by the two other approaches) and the structural analysis of environmental public goods.

These three approaches conflict with one another on several issues, partly because they are dealing with different dimensions of environmental politics, respectively the preconditions of cooperation, methods of decision-making, and a just framework. Nonetheless, despite their divergence, they all share a common feature, namely a commitment to the idea of neutrality, respectively derived from objective science, value-neutral economics, and the idea of political neutrality. While valuable, such neutrality has its limits as a normative basis for public policy and in particular for environmental policy. Examining these limits, primarily through internal criticism of the different approaches, is the focus of this dissertation. Identifying these limits creates a general framework for dealing with environmental public goods. While broad, such a framework bridges the current gap between so-called ‘green political theory’ and mainstream political and economic theory.

REFERENCES

Gardiner, Stephen M. 2011. *A perfect moral storm: the ethical tragedy of climate change*. New York: Oxford University Press.

Stijn Neuteleers obtained his PhD in philosophy from KU Leuven, Belgium. He was supervised by Antoon Vandevelde (Institute of Philosophy). He is currently a postdoctoral researcher at the Research

Centre for Environmental Studies and Sustainability (CREIDD) at the
Université de Technologie de Troyes (France).

Contact e-mail: <stijn.neuteleers@gmail.com>

Website: <<http://utt.academia.edu/StijnNeuteleers/>>