ABSTRACT

The comments on Dirk Helbing and my paper “Homo Socialis: An Analytical Core for Sociological Theory” have provided many insightful suggestions. Several, such as Siegwart Lindenberg’s (2015) proposal to include flexible activation to the core and David Wolpert’s (2015) persona model, I take to be interesting complements to our suggestions. Others, such as Michael Macy’s (2015), include admirable interpretations of our argument. I will here address a subset of issues where commentators have questioned some aspects of our proposal. I thank the commentators for helping me think more carefully about some key issues.

1 The Rational Actor Model

The rational actor model treats individuals as having a preference function over payoffs, having access to various actions that affect the probability distribution over these payoffs, and having beliefs concerning the relationship between actions and payoffs. Following Savage (1954), we defined a rational actor as one who exhibits consistent choices (transitivity), whose preference for action A over action B does not depend on other available choices (independence from irrelevant alternatives), and who does not engage in “wishful thinking” (the probabilities assigned to various actions do not depend on the preferences over the payoffs). We observed that a rational actor can be modeled as maximizing a preference function over the set of available choices.

We stressed that preferences, payoffs, and beliefs can all be time-, state-, and social-context dependent, and we showed how to treat cases where the stochastic nature of imperfectly measured state variables may lead an individual to make distinct choices under what appear to be identical circumstances, following the constructions of Luce and Suppes (1965) and McFadden (1973). We noted
that the resulting construction is extremely powerful, being the basis for core analytical models in contemporary biology and economics. Moreover, accepting the rational actor model is a prerequisite for using game theoretic reasoning to explain behavior involving strategic interaction. Yet the rational actor model is a rather weak construct because it does not specify the content of preferences or beliefs, and it does not assert that an agent’s choices necessarily reflect the agent’s objective physiological or psychological needs, and hence their satisfaction does not necessarily improve the agent’s well-being. Indeed, a rational actor can consistently wish he had a different preference ordering (Gintis, 1972).

1.1 Claim: Rational Choice Theory Cannot be Falsified

“Because utility is unobservable,” writes Geoffrey Hodgson, “all kinds of behavior can be “expressed” in terms of utility maximization, without fear of refutation . . . no evidence can possibly refute the theory that agents are maximizing some hidden or unknown variable (such as utility).”

In fact, if any one of the Savage (1954) axioms fail, the rational actor model will be falsified for the situation in question. For instance, suppose we observe an individual making pairwise choices from a set of \( n \) payoffs. The number of choice pairs is then \( m = n(n - 1)/2 \), so in general the individual’s behavior can exhibit \( 2^m \) distinct choice patterns. However, the transitivity axiom in rational choice theory allows only \( n! = n(n - 1) \cdots 2 \cdot 1 \) choice patterns. For instance, if \( n = 10 \), then a direct calculation shows that only about one choice pattern in ten million is compatible with utility maximization. We thus see that, given a small choice set, the assumption of utility maximization can easily be refuted when it is false (Andreoni and Miller, 2002).

Hodgson perhaps means that no evidence can possibly refute the theory that there is some unknown choice space over which the individual’s choices can be represented by a preference function. This statement is correct, but, in fact the rational actor model is justified by its usefulness, not by its universal applicability. Moreover, in many salient cases of seemingly non-rational behavior fits the rational actor model, when the choice space is appropriately chosen. My book The Bounds of Reason (2009b) analyzes prominent examples where the proper specification of a choice space transforms preference inconsistency into conformity with the rational actor model. One of these is adding the status quo point to a choice model involving possible gains and losses. This permits the treatment of prospect theory within the rational actor model (Kahneman and Tversky, 1979). Another is including the time of choice in the choice space to handle time inconsistency problems associated with non-exponential discounting (Ahlbrecht and Weber, 1995). The goal here is not to rescue rational choice theory, because we recognize that it is not universally applicable, but rather to apply it to novel problems through the appropriate specification of the choice space.
In the same vein, Hodgson asserts: “A key problem with utility maximization is that it is so general that it can fit anything.” I think it is more reasonable to say that the rational actor model fits any form of decision-making behavior in which the Savage axioms are satisfied. Like any mathematical structure, the rational actor model is of little use supplemented by specific content — in this case, a more detailed specification of the preferences and beliefs of the decision-maker.

1.2 Claim: The Rational Actor Model Obscures Dynamic Mental Life

In reaction to an earlier draft of these remarks, Vernon Smith (personal communication) writes that Adam Smith’s *Theory of Moral Sentiments* does not reduce to [the rational actor model] subject to social restraints; … Its a process, a maturation process of learning to see ourselves as others see us, engage in the pleasures of mutual empathy, which is not utilitarian, but concerns synchrony and fellow-feeling. Discordance or disharmony follows from a disconnect between the individuals rule-following behavior and the emergent/emerging social conduct rule. And it was done in 1759/1790 and it is fresh and relevant for the 21st century. Modern equilibrium theory is still plagued by its efforts to jump over process, to avoid thinking through social experience, the emotions of gratitude and resentment that form sociability.

Vernon Smith’s graduate economics education at Harvard was probably not much different than mine in dealing with the rational actor model. I learned that rational choice is a theory of human behavior, including having self-interested goals, optimally choosing instrumental means toward achieving these goals (Max Weber’s *Zweckrationalität*) and unlimited costless information processing. I suspect that graduate students today are still learning the same things in classrooms around the world, despite the fact that the formal theory, fully developed more than sixty years ago, says no such thing at all. Moreover, behavioral scientists in other disciplines often follow this informal understanding of the rationality postulate. Thus, I fully agree with Vernon Smith’s remarks as regards the informal socialization of economists, but this does not undermine the rational actor model itself. A short history of the modern concept of rational actor is perhaps in order.

Paul Samuelson (1938) recognized that all that is needed for standard microeconomics is a severely limited concept of preference consistency, which was extended by John von Neumann and Oskar Morgenstern (1944) to choice under uncertainty, popularized in Samuelson’s pathbreaking *Foundations* (1947), and fully axiomatized shortly thereafter by Leonard Savage (1954). John Maynard Smith (1982), John Alcock (1993), and others subsequently
deployed the same limited concept of rationality to explain animal behavior. I have argued that rational choice theory in this limited form is a building block for all the behavioral sciences (Gintis, 2007b, 2009). It is this elegant and powerful mathematical construct that Dirk Helbing and I present as a building block of sociological theory.

Rational choice theory, therefore is not a theory of behavior, but rather a mathematical construct, like a vector space or a differentiable manifold, with very strong general properties that can be deployed wherever the appropriate consistency conditions apply.

I agree with Hodgson that “On its own, rational choice theory tells us nothing of significance that is specific to the human psyche, human interaction, human nature, or human society.” Hodgson considers this a weakness, but in view of the capacity of rational choice theory to reduce choice models to those that maximize a preference function subject to constraints, this theory is a powerful analytical basis for specific theories of human behavior. We may not always want to make use of this analytical framework in dealing with a particular form of behavior. Vernon Smith (2015) suggests, for instance, that the rational actor model points us away from the true mainsprings of human behavior. But, unless this behavior is substantively irrational, it is always available to us and can complement other more insightful ways of modeling behavior. For instance, Alan Isaac (2015) suggests that “role theory often provides a plausible substitute for the rational actor model.” But role theory does not assert that behavior is irrational, and thus is complement to, not a substitute for, rational choice. This is precisely how we treated role theory in our paper.

As Robert Goldstone (2015) writes:

> Constraining agent behavior by rationality is an effective way to generate strong predictions for patterns that can be observed at the collective level, but is by no means the only one. Other possible sources of constraint are evolution, neuroscience, and individual cognition.

The error of asserting that rationality is sufficient to explain human behavior is a special form of methodological individualism exhibited in the past by many economic theorists. I do not share any such notion. Considerable material concerning human behavior in addition to the rational actor model should be included in the analytical core.

Prior to the maturation of experimental and behavioral game theory, a legitimate critique of attempts to explain behavior violating the rational material self-interest assumption by positing “exotic” non-self-regarding preferences was that such explanations were not empirically testable and hence vacuous — like trying to explain biological life by positing an ineffable élán vital. However, we are now capable of adding substance to the bare form of rational choice theory
through laboratory and field testing of alternative behavioral models. The capacity to explore empirically the content of preference orderings considerably expands the power of the rational actor model.

1.3 Claim: The Rational Actor Model Obscures Human Action

“Because of the neo-classical outcome-payoff focus,” asserts Vernon Smith (2015), “Max-U [i.e., the rational *Homo economicus* actor] failed to define a natural on-the-ground treatment of rule-governed human sociality as in TMS [Adam Smith’s Theory of Natural sentiments] . . . Even if a utilitarian route [such as that offered by Gintis and Helbing] can be shown to possess technical equivalence with the TMS model, it fails to convey an understanding of the mainsprings of human action. Smith fully recognized that the emergent rules of his *Homo socialis* model might well be efficient, but warned that this corollary does not enable us to understand the process whereby these rules came about.

This observation is doubtless true, but it does not diminish the value of the rational actor model, the analytical power of which does not depend on a deep understanding of the human mind. However, the rational actor model can help us uncover deep workings of the human mind. Using the rational actor model in experimental game theory has led us to many insights that were previously unavailable (Gintis *et al.*, 2005; Gintis, 2009b). Working out the physiology and psychology of the mind have taken us into areas of research far removed from abstract decision theory, but except in certain limited and well-specified areas, they have not revealed mechanisms that undermine the principles of rational choice theory.

1.4 Claim: Morality Cannot be Represented by a Preference Function

“[M]orality, altruism and other-regarding preferences are different things,” maintains Geoffrey Hodgson (2015). “Drawing on what I believe is the majority view among moral philosophers,” he continues, “I maintain that morality cannot be adequately summarized by any preference function.” Exploring the views of prominent moral philosophers, Hodgson concludes with John Mackie that morality “is absolute, not contingent upon any desire or preference or policy or choice.”

In our essay, as in virtually all of the contemporary research on human social behavior, morality is an expression of how people behave, both in what they assert and in the choices they make. We may call this behavioral morality. The content of morality as developed in the works of moral philosophers and theologians we may call abstract morality. It is an extremely important question as to when behavioral morality conforms to the canons of abstract morality (Frohlich and Oppenheimer, 1992, Rozin *et al.*, 1999, Greene and Haidt, 2002), and there is little doubt but that there is a strong relationship
between the two (Brown, 1991). However, it is clear than even in the realm of abstract morality, when individuals must decide on a course of action, they will generally face a variety of moral trade-offs without a plausible manner of placing them in lexicographic descending order. Trade-offs are therefore inevitable. Moreover, any abstract morality that maintains that there is some moral principle that always trumps the preferences of individual actors is likely to be false. Moreover, many of the most important moral principles are supererogatory (Heyd, 2012), meaning they are worthy but not obligatory. This includes giving to charity, helping strangers, gossiping about the good deeds and misdeeds of associates, rewarding the moral acts of others, and punishing violators of social norms. In the case of supererogatory acts, the trade-off between morality and personal preference becomes ineluctably paramount.

Moving to behavioral morality, the notion that such morality does not involve trade-offs life is demonstrably false. Take, for instance, the insightful experiment of Gneezy (2005), who studied 450 undergraduate participants paired off to play three games, in which all payoffs were of the form \((b, a)\), where player 1, Bob, receives \(b\) and player 2, Alice, receives \(a\). In all games, Bob was shown two pairs of payoffs, \(A: (x, y)\) and \(B: (z, w)\) where \(x, y, z, \) and \(w\) are amounts of money with \(x < z\) and \(y > w\), so in all cases \(B\) is better for Bob and \(A\) is better for Alice. Bob could then say to Alice, who could not see the amounts of money, either “Option \(A\) will earn you more money than option \(B\),” or “Option \(B\) will earn you more money than option \(A\).” The first game was \(A: (5, 6)\) vs. \(B: (6, 5)\) so Bob could gain 1 by lying and being believed while imposing a cost of 1 on Alice. The second game was \(A: (5, 15)\) vs. \(B: (6, 5)\), so Bob could gain 1 by lying and being believed, while still imposing a cost of 10 on Alice. The third game was \(A: (5, 15)\) vs. \(B: (15, 5)\), so Bob could gain 10 by lying and being believed, while imposing a cost of 10 on Alice.

Before starting play, Gneezy asked the various Bobs whether they expected their advice to be followed. He induced honest responses by promising to reward subjects whose guesses were correct. He found that 82% of Bobs expected their advice to be followed (the actual number was 78%). It follows from the Bobs’ expectations that if they were self-regarding, they would always lie and recommend \(B\) to Alice.

The experimenters found that, in the second game, where lying was very costly to Alice and the gain from lying was small for Bob, only 17% of Bobs lied. In game 1, where the cost of lying to Alice was only 1, but the gain to Bob was the same as in game 2, 36% of Bobs lied. In other words, Bobs were loath to lie but considerably more so when it was costly to Alices. In the third game, where the gain from lying was large for Bob and equal to the loss to Alice, fully 52% of Bobs lied.

This shows that many subjects are willing to sacrifice material gain to avoid lying in a one-shot anonymous interaction, their willingness to lie increasing
with an increased cost to them of truth telling, and decreasing with an increased cost to their partners of being deceived. Similar results were found by Boles et al. (2000) and Charness and Dufwenberg (2006). Gunnthorsdottir et al. (2002) and Burks et al. (2003) have shown that a socio-psychological measure of “Machiavellianism” predicts which subjects are likely to be trustworthy and trusting. Batson et al. (1997) documents the self-delusion involved in affirming a moral principle and yet at the same time violating it.

1.5 Claim: Homo Socialis has no Need for Social Preferences

“The hypothesis that social preference can describe *Homo socialis*,” writes Vernon Smith (2015), “reduces inevitably to a rescue of neo-classical economics in which Max-U(own payoff, other payoff) substitutes for Max-U(own payoff).” He concludes that “*Homo socialis* has no need for the artifice of utilitarian social preferences.”

Yet, there are clear cases where non-self-regarding behavior is insightfully modeled by assuming an actor trades off payoffs to himself against payoffs to a transaction partner. For instance, the tendency for responders in the ultimatum game to reject low positive offers can be explained by their intention of punishing proposers for unfair behavior. In a variant of the game in which offers are generated by a computer rather than by the proposer, and if responders know this, low offers are rarely rejected (Blount, 1995). This suggests that players in the standard prisoners dilemma are motivated not simply by the violation of a fairness rule, but rather by the goal of punishing the proposer. Moreover, in a variant of the game in which a responder rejection leads to the responder getting nothing but allows the proposer to keep the share he suggested for himself, responders almost never reject offers, and proposers make considerably smaller (but still positive) offers (Bolton and Zwick, 1995). A plausible explanation of this behavior is that when the responder cannot hurt the proposer of an unfair offer, he will accept the unfair offer.

One might argue that ultimatum games are not important outside the laboratory. Consider, however the everyday life importance of *fair pricing*, where a consumer will refuse to purchase a product, because he feels it is overpriced, even though he values the product more than its purchase price, and in other circumstances he would happily pay that price. My experience is that such behavior is quite pervasive in our society. In some cases, this behavior can be explained as rational search behavior, but very often it cannot, because the value of the product will disappear before another opportunity to purchase arrives. For instance, in a laboratory experiment described in Thaler (1985), the following story is presented to half of a group of experimental subjects:

You are lying on the beach on a hot day. All you have to drink is ice water. For the last hour you have been thinking about how
much you would enjoy a nice cold bottle of your favorite brand of beer. A companion gets up to go make a phone call and offers to bring back a beer from the only nearby place where beer is sold, a fancy resort hotel. He says that the beer might be expensive and so asks how much you are willing to pay for the beer. He says that he will buy the beer if it costs as much or less than the price you state. But, if it costs more than the price you state he will not buy it. You trust your friend, and there is no possibility of bargaining with the bartender. What price do you tell him?

The other half of the subject pool is shown the same text, except that “a fancy resort hotel” is replaced by “a small, run-down grocery store” and “bartender” is replaced by “store owner.” Thaler (1985) reports that the median responses for the two versions were $2.65 for the resort hotel and $1.50 for the grocery store. This, of course, is a comparison of two ultimatum games, and the fact that a subject values a beer more than $2.64 but refuses it from the grocery store at $1.50 is an archetypal rejection of a positive offer in the ultimatum game. Ultimatum games of this sort are extremely common in daily life.

Nevertheless, Vernon Smith is correct in saying that not all, and perhaps not even most social behavior can be explained by including the payoff to transaction partners in an actor’s preference function. We were very explicit on this point by including in other-regarding preferences a taste for cooperation, fairness and retribution, the ability to value such character virtues as honesty, hard work, piety and loyalty, and the ubiquitous susceptibility to caring about the esteem of others. One is honest, for instance, not because one cares about one’s transaction partners, but because honesty is a virtue. An honest person will behave honestly even when this is materially costly and benefits others for whom one has no or even negative regard. Sugden, 1982, Andreoni (1990, 1995), and others have similarly shown that charitable giving cannot be represented by adding a term for the utility to the recipient in the donor’s preference function.

1.6 Claim: Sociality is Rule-Following, not Social Preferences

“Human sociality,” observes Vernon Smith (2015), “is expressed in rule following conduct, where a rule involves choosing an action based on the actor’s judgment of its propriety, given its context (or circumstances), where context includes the alternative outcome payoffs in the choice set.” Following Adam Smith, Vernon Smith notes that appropriate behavior in a given context involves following specific rules associated with this context rather than a utility function that reflects the actor’s context-independent preferences. Models in our paper are perfectly consonant with this important insight. We explicitly say that individual preferences are “time, state, and social context dependent.”
Adam Smith’s *Theory of Moral Sentiments* (1759/2000), on which Vernon Smith draws for his treatment of rule-following, is in fact, among many other things, a prescient precursor of the modern role-actor model developed by Ralph Linton (1936) and Talcott Parsons (1937). Vernon Smith’s rule-following framework is precisely that of our model of general social equilibrium, where we say:

Actors are in general rational decisions-makers who maximize their preference functions subject to the content of the social roles they occupy, and given a belief system that is context-dependent and governed by the expectations defined by the actor’s social location. These decisions determine the social actors’ role-specific behaviors.

Interestingly, we follow this statement with the example of engaging a taxi in a strange city, which is much like Vernon Smith’s example of the Korean baggage handlers and Lindenberg’s (2015) flexible activation scenarios.

When framed in terms of social role theory and the psycho-social theory of norms, there is simply no conflict between the rational actor model and Vernon Smith’s (2015) treatment of sociality as rule-following behavior.

## 2 General Social Equilibrium

Our general social equilibrium model treats individuals as assuming a variety of social roles, each of which includes conventional and normative rules for appropriate behavior and incorporates the material incentives and normative principles on which role performance is based. This model draws upon the rational actor model to represent the way individuals choose among the roles they will assume and the actions they take within their assumed roles. In many cases, social roles are organized by institutions and organizations, just as firms organize the division of labor in general economic equilibrium, although institutional goals usually go beyond monetary gain and the role of moral incentives may be of predominante importance outside the economic sphere.\(^1\)

Our notion of general social equilibrium codifies the ideas of many social thinkers who have addressed the problem of how society regulates the interaction of individuals in a manner, that permits the successful reproduction of highly complex social relationships from period to period. Several commentators have expressed problems with our treatment.

---

\(^1\)In fact, economic firms, even under competitive conditions, can implement non-monetary goals and yet survive and prosper if their owners, clients, and/or employees are willing to sacrifice monetary gain for these non-monetary goals. This is sometimes referred to as *social entrepreneurship* (Bornstein, 2007; Gintis 2008).
2.1 Claim: General Social Equilibrium Cannot Model Conflict

The term *equilibrium* conveys to many a sense of calm, order, and stasis, that is quite foreign to the use of the term in dynamical systems theory. It is for this reason, says Paul Lewis (2015), that Hayek chose to avoid the term.

In the 1960s, Hayek explicitly abandoned the notion of equilibrium in favour of the notion of ‘order,’ writing that: “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 been discovered and competition has therefore ceased.”

Alan Isaac (2015) similarly expresses the sentiments of many in saying that Dirk Helbing and I “consider norm conflict a disequilibrium phenomenon rather than a normal state of affairs.” We retain the term *equilibrium*, but we do not mean it to convey harmony or stasis. However, general social equilibrium does not imply social harmony. Norm conflict is quite compatible with general social equilibrium. Social equilibrium is a situation in which, at each point in time, no individual, given the constraints he faces, has an incentive to change his choice of social roles to fill or his behavior within these roles, given the current society-wide distribution of roles, the aggregate assignment of individuals to roles, and the macrosocial pattern of role incentives and social norms.

This situation does not preclude norm conflict. General social equilibrium can insightfully model crime, corruption, and other social pathologies and conflict scenarios that have chronic and continuing social presence.

Conflict over the rules of the game do indeed lie outside the general economic equilibrium model. Ulrich Witt writes:

> free markets... foster the simultaneous attainment of individual optima — an unplanned harmony that emerges spontaneously as if guided by an invisible hand. Can this thought experiment be generalized...[to] an unplanned, spontaneously emerging harmony of social interactions in a general sense? It cannot, I think, simply because the (idealized) conditions of the price mechanism do not apply to social interactions in general.

This is true. However, it is not the nature of the price system that prevents the general economic equilibrium model from recognizing conflict. It is rather its reduction of all social relationships to exchange relationships based on given prices. Prices, being given, are being inherently non-conflictual. The general social equilibrium model does not suffer from this limitation.

General social equilibrium therefore does not suggest social order or efficiency. The same, of course, is true of general economic equilibrium. Supporters of *laissez faire* often assert that free market exchange is socially efficient and

---

2A somewhat more general definition of general social equilibrium is that the fraction of agents occupying any given social role is constant over time.
conflict-free (von Mises, 2007), and indeed the First Welfare Theorem of general economic equilibrium asserts that under certain conditions market equilibrium is socially efficient (Mas-Colell et al., 1995). But those conditions are systematically violated in all known economies (Musgrave, 1959; Atkinson and Stiglitz, 1980). No serious economist believes that real-world market economies satisfy the conditions for optimality, although many economists believe that market imperfections are often less harmful than the governmental instituted to correct them (Friedman, 1962; Buchanan et al., 1980). General economic equilibrium, like its social generalization, does not imply socially optimal behavior.

2.2 Claim: Disequilibrium is the Normal State of Affairs

Ulrich Witt writes:

Can [general equilibrium theory] hold promise for the discipline of sociology when, even in the domain of economics, evidence supporting the claim that all agents can realize their individual optima simultaneously, i.e. that a state of general equilibrium is empirically observable in present-day reality, is lacking?

The claim that disequilibrium is the normal state of affairs is obviously correct. Just a general social equilibrium is not necessarily efficient or conflict-free, neither is likely ever to be attained in any real-world society. The same is of course true of the far simpler general economic equilibrium model.

Of course, an equilibrium model is useless, unless there is some expectation that an equilibrium state is in some sense a focal point for the dynamics of the system under study. The most favorable sense is that the equilibrium is locally stable in an appropriate dynamic, and the basin of attraction of the equilibrium is sufficiently large that escapes from the basin of attraction through repeated shocks, are rare (Hirsch and Smale, 1974; Hofbauer and Sigmund, 1998; Staddon, 2001). Other possibilities are that the dynamical system move randomly among the basins of attraction of two or more locally stable equilibria (Freidlin and Wentzell, 1984), or the equilibrium point is a strange attractor of a chaotic dynamical system (Ruelle, 1989).

We are unfortunately unable to specify a plausible social dynamics for which our general social equilibria are the stationary states. There is little doubt that such dynamics exist, because societies generally alter their constellation of social roles and agent behaviors rather slowly in response to changes in technology and the environment, although many societies also experience infrequent rapid shifts in social organization. We are currently incapable of outlining the nature of such dynamics in any analytical expressible form.

In economics, general equilibrium theory has been in a similar situation. Walras knew that his model needed a theory of fully decentralized price adjustment that would ensure stability of equilibrium. He never offered such a
theory, but rather proposed a "tâtonnement" process in which a centralized agent, the so-called auctioneer, adjusts prices until equilibrium is attained, only after which are the economic actors permitted to produce, exchange, consume, and invest.

The stability of the Walrasian economy became a central research focus in the years following the existence proofs (Arrow and Hurwicz, 1958, 1959, 1960; Arrow et al., 1959; Naiako, 1959; McKenzie, 1960; Naiako and Uzawa, 1960). Following Walras’ centralized tâtonnement process, these models assumed that the production and trade do not take place until equilibrium prices are announced, and out of equilibrium, there is a price profile shared by all agents, the time rate of change of which is a function of excess demand. These efforts at proving stability were successful only by assuming narrow and implausible conditions without microeconomic foundations (Fisher, 1983). Indeed, Scarf (1960) and Gale (1963) provided simple examples of unstable Walrasian equilibria under a tâtonnement dynamic, and more recent researchers have shown that, no general stability theorem could be obtained based on Walras’ tâtonnement process. Subsequent analysis showed that chaotic price movements are the generic case for the tâtonnement adjustment processes (Saari, 1985; Bala and Majumdar, 1992).

Despite the weakness of having no known attractive stability properties, the general equilibrium model has remained part of the core of modern economic theory. We proposed that an enriched general social equilibrium model similarly become part of the core of modern sociological theory. In the case of economic theory, we have moved a long way towards an acceptable approach to market dynamics. New directions in economic dynamics model the economy as an evolutionary game (Taylor and Jonker, 1978; Friedman, 1991; Weibull, 1995; Gintis, 2009a) in which the actors’ strategies consist of plans for production, consumption, and work based on their personal assessment of the contracts to which they are willing to agree (Gintis, 2007a). These models are finite Markov processes with huge state spaces. We can analytically determine that the stable Nash equilibria of the resulting evolutionary game are precisely the market-clearing solutions of the general equilibrium model (Gintis and Mandel, 2012, 2014). But the state space of the Markov process is far too large to determine its stationary distribution analytically, so we are forced to sample its behavior through computer modeling (Gintis, 2013). We find that, even for very different specifications of the institutional structure and production conditions, the market-clearing equilibria are always stable (Gintis, 2012), although in an economy with complex institutions, the economy can exhibit great volatility even in the absence of systemic shocks (Gintis, 2007a).

The reason for the success of these models, despite the generally chaotic movement of prices in traditional tâtonnement models, is that instead of a common set of prices faced by all agents, our evolutionary dynamical models assume that each actor has his own estimate of the price structure. The actor
uses these private prices as a basis for his choices, and he updates his price structure by copying more successful agents. Only in equilibrium does the resulting pattern of heterogeneous prices converge to something resembling uniform prices for all agents. No researcher has attempted to extend these results to a general social equilibrium setting, but there is no reason in principle that it could not be done.

It might be thought that a general social equilibrium model would need much more complex adjustment processes that are needed in general economic equilibrium, where only prices need adjust. However, even in general economic equilibrium, as prices adjust, so do all the activities of producers and consumers, so the social adjustment process is not qualitatively different from the economic adjustment process.

2.3 Claim: The General Social Equilibrium Model is Unrealistic

Models are unrealistic by construction — they abstract from certain real-world elements in order to focus upon other elements that are postulated to be important for understanding certain real-world behaviors. The goal of model-building to create a tractable analytical structure, analyze the behavior of this structure, and test the fruitfulness of the results by comparing them with empirical data. The General Social Equilibrium model is appropriate for expressing some major articulation points between individuals and social roles, for depicting the contribution of social conventions and social norms in coordinating complex social interactions, for modeling how important social interactions depend on standard social cues and cultural expectations, and for illuminating many other standard sociological phenomena. The model cannot handle any macrodynamical phenomena (for the time being) and hence cannot deal insightfully with the social processes involved in fundamental cultural and institutional change.

For an example of the usefulness of the general economic equilibrium model in its most abstract form, despite its “unreality,” I submit that it lays the foundation for contemporary public economics, the theory of the conditions under which free markets fail to promote allocative efficiency and how regulatory interventions might correct such failures. This theory was perfected in the early post World War II period, and remain valid today, more than a half century later. The theory recognizes several types of market failure, including:

a. Increasing Returns to Scale: In some sectors, the efficient firm size is so large that competition is precluded. For instance, water to a city may

---

3This interpretation of model-building was espoused by Milton Friedman (1953). Its contemporary proponents include Godfrey-Smith (2006, 2009), Weisberg (2007), and Wimsatt (2007, Ch. 6).
be supplied by a single reservoir and a unified system of processing and delivery. There is simply no room for multiple firms to compete without degrading the system’s efficiency.

b. Public Goods: Some goods are non-exclusive — they are consumed equally by many or all individuals (although they may be valued differently by different individuals). For instance, national defense protects all equally, and many forms of public health measures affect the incidence of diseases for the entire population.

c. Externalities: Some goods are produced using technologies that release waste products into the environment at zero or low cost to the producer but that impose high costs on others.

d. Unverifiable Product Quality: There are quality good industries, in which the quality of a product cannot be ascertained until after purchase, and in which reputation effects are not sufficient to ensure an appropriate quality level.

While there are serious theoretical problems with public economics (Lipsey and Lancaster, 1956–1957), I know of no serious problems with the applicability of public economic theory, nor of any plausible alternative thereto. The “unreality” of general economic equilibrium does not diminish its usefulness.

2.4 Claim: General Equilibrium Macroeconomics is a Failure

Paul Ormerod (2015) notes that modern “rational expectations” macroeconomic theory, as developed by Robert Lucas and his coworkers (Lucas, 1981; Stokey et al., 1989) has turned out to be a disappointment. He writes:

[T]he attempts in recent decades to put macroeconomic theory on [general equilibrium] micro foundations, initially in real business cycle theory, and then more recently in dynamic stochastic general equilibrium theory (DSGE) . . . encouraged economists to promote an intellectual climate in which both regulators and policy maker came to believe that the problems of business cycles had been eliminated. Lucas (2003) claimed that “the central problem of depression-prevention [has] been solved, for all practical purposes.”

I agree that this theory has been weak, but I do not agree that rational expectations macroeconomics is based on general equilibrium theory. The claim that contemporary macroeconomic theory is based on the general equilibrium model, has no substance because as we explained above, there has been no general dynamic market model until recently, and recent dynamic models have little relationship to standard macroeconomic theory.
The term *rational expectations* expresses the notion that one could expect rational economic agents to understand the effects of government fiscal and monetary policy at least as fully as professional economists based on a full macroeconomic model. The term was appropriate as a criticism of the Keynesian unemployment rate vs. inflation trade-off schedule (the so-called *Phillips curve*), which can exist only if agents are continually fooled concerning the level of next-period prices. But in a valid general equilibrium model, the only rational expectations are adaptive expectations, and such expectations do not require knowing “the macroeconomic model,” but rather in choosing among the successful strategies exhibited in the current period (see the paper by Gallegati and Palestrini referred to in the commentary of Mauro Gallegati, 2015). Indeed, to the extent that the economy is a complex dynamical system, there is only a limited range over which agents can reasonable predict the effects of policy decisions or the introduction of new technologies.

Rational expectations macroeconomics purports to be solidly based on widely-accepted microeconomic principles that accurately describe the market economy (the “rational” in rational expectations), but this claim is untenable. The market economy is in fact a complex dynamic system whose behavior can be simulated, much as the weather is simulated by supercomputers, but cannot be captured in a few recursive equations, as the rational expectations macroeconomics supporters claim. Instead of a large number of economic actors, rational expectations macroeconomics assumes there is one “representative agent,” and instead of large numbers of firms and industries, rational expectations macroeconomics assumes there is one “production function.” The only sources of volatility in such a world are technology shocks and the vagaries of “expectations,” which are subjective and difficult to measure (Kirman, 1989).

Among the more bizarre modeling choices of rational expectations macroeconomics is to assume that all markets clear instantaneously. This, of course assumes away the coordination failures that underlie macroeconomic fluctuations, and is quite uncharacteristic of general equilibrium theory. This sort of macroeconomic modeling would be acceptable, if the resulting models predicted well, but they do not (Frydman and Goldberg, 2011).

### 2.5 Claim: The General Equilibrium Model Cannot Model Power

Ulrich Witt suggests that general equilibrium abstracts from power relations and thereby cannot represent a central characteristic of real societies. He writes:

> [F]reedom of contract... underlying general economic equilibrium theory... may well be an acceptable approximation of the actual conditions under which transactions take place in modern markets under the rule of law. In contrast, assuming analogously freedom to interact as a universal condition in the social domain would
by no means correspond to the real-life conditions with which sociology is occupied, even in modern, liberal societies. [T]he crucial sociological category of (formal and informal) power suggests a quite different, and less harmonious, idea of what represents the analytical core of sociological theory.

In fact, Samuel Bowles and I developed a model of political power in the economy based on rejecting just one assumption of the general economic equilibrium model. This is the assumption that all exchanges are enforceable (e.g., by the judicial system) at no cost to the exchanging parties. We argue (Bowles and Gintis, 1990) that where some aspect of the object of exchange is so complex or difficult to monitor that comprehensive contracts are not feasible or enforceable by a third party, exogenous claim enforcement cannot obtain. We also argued that endogenous enforcement is quite general, including some of the most salient exchanges in a market economy — labor, financial, and many consumer goods markets.

In the typical labor market, the employer pays a wage in return for the worker’s agreeing to come to work certain days and hours of the week. The intensity, care, and creativity, the worker bestows upon his tasks during these hours is not specified in the contract between the employer and employee, and cannot be enforced by a judicial authority. The worker’s performance must be evaluated by the firm’s internal system of surveillance, accountability, and authority. This system involves power relations, in the following sense: for A to have power over B, it is sufficient that, by imposing or threatening to impose sanctions on B, A is capable of affecting B’s actions in ways that further A’s interests while B lacks this capacity with respect to A. This holds in the firm where A is the employer and B is the employee. We also show in this and related papers that this also holds where A is a lender and B is a borrower, and where A is the consumer of a quality good and B is a firm producing this good.

In the labor market, the power of employer over employee consists in the firm’s ability to dismiss or otherwise impose sanctions on the worker. In a financial market, the power of the lender is to withdraw a line of credit or compromise the borrower’s credit rating. In a quality good market, the consumer’s power over the producer is to switch allegiance to another supplier (Gintis, 1989).

This power exists in labor markets only if there is a positive unemployment rate in equilibrium; in financial markets only if there is credit rationing (the demand for loans exceeds the supply in equilibrium); and in quality good markets only if firms are quantity constrained (they would like to increase their sales) in equilibrium. We call this sort of economic power short-side power, because the agent with power is the agent on the short side of the market — the employer in labor markets, the lender in financial markets, and the consumer in quality good markets.
We have not extended this notion of socially endogenous power to general social equilibrium in our paper, but there is no reason our model of economic short-side power could not be extended to cover non-economic relations. Moreover, in a general social equilibrium model, even purely coercive power relations can be directly modeled, as we there allow for social institutions that operate outside the context of voluntary agreements.

2.6 Claim: General Social Equilibrium is Individualistic

Ulrich Witt writes:

Many sociologists would... [oppose game theoretic formulations of social life], on account of its individualistic foundation of sociological theory. Such a foundation has been debated in sociology at least since Homans (1961) [Social Behavior: Its Elementary Forms], and has not found much support in the discipline.

Alan Isaac (2015) similarly notes:

Given the emphasis that they [Gintis and Helbing] give to individual rational choice and evolutionary game theory, they seem to have in mind primarily that the adaptation is at the level of individual humans. Some sociologists will find this rather like an anatomist modeling a body as a collection of cells rather than as a collection of organs: not exactly incorrect, but rather at odds with the subject matter.

Methodological individualism is usually understood as a doctrine holding that social life can be understood in terms of the preference, beliefs, and actions of individual social actors. This tenet has often been asserted, but it certainly has never been proven. My book Bounds of Reason (2009b) is an extended attack on methodological individualism, and Dirk Helbing and my contribution here is certainly not methodologically individualist. It is commendable that methodological individuals have not found much support among sociologists.

Our analytical core is, however, individualist in the restricted sense of viewing social stability and change as resulting from an interaction between individuals on the one hand and social structures on the other, the latter including conventions, norms, and the network of social institutions that regulate the structure of social roles. There are of course social theorists who argue that there is no role for individuals as fundamental units in social theory because individuals are determined by larger social aggregates. As Michael Hechter (1987) argues, these theorists claim much, but deliver little. The notion that any area of social theory could successfully model social processes
without modeling individual behavior from the point of view of that the actors’ motivations and psychological processes is implausible. Our ‘individualism’ is therefore justified.

3 Social Theory and the Analytical Core

Alan Isaac (2015) asks:

Although most economists are comfortable with the identification of theory with mathematics, this is much less the case in sociology. How many of the questions central to sociological theory seem amenable to mathematicization? Should those that are not amenable be obliged to reference an analytical core? To provide a specific example, do the authors believe that the concept of a lifeworld is either unimportant or mathematizable? If neither, what should be the relationship of this central concept in sociological theory to their proposed analytical core?

We did not claim that our suggested analytical core is complete, and additions to the analytical core may well be non-mathematical. The notion that all sociological theory is mathematical or can be framed in mathematical models is implausible. Non-mathematical contributions to sociological theory may build on empirically validated mathematical models, but this need not be the case.

Moreover models, mathematical or otherwise, are merely tools in the social theorist’s tool box. Because human beings are complex adaptive biological creatures and societies are complex adaptive systems, neither is likely ever to be adequately described by a fixed set of analytical models. As Peter Godfrey-Smith (2006) and William Wimsatt (2007, Ch. 6) have observed, a model isolates a few causal processes and relationships that are believed to be central to explaining a phenomenon, derives the logical implications of their operation in the absence of other influences, and measures the success of the model by its correspondence with the observed phenomenon.

For instance, the rational actor model is a powerful analytical device, but it does not account for all human choice behavior (Gintis, 2007b, Ch. 1). In the Allais paradox mentioned by David Wolpert (2015), and in the even more important Ellsberg paradox, choice behavior violate the Savage axioms. Moreover, there are known systematic deviations from the linearity of probabilities in the expression of expected utility that occur when one or more of the options has an extremely high (positive or negative) payoff with extremely small probability (Machina, 1989). The violation of the no wishful thinking axiom, mentioned both by Robert Goldstone (2015) and Michael Hechter (2015), is often important, as when individuals overvalue evidence that supports their
beliefs and undervalue evidence that undermines these beliefs (Kahan et al., 2005). I do not know the implications of dropping the no wishful thinking assumption, but I suspect it implies that probabilities do not enter linearly into the expression for expected utility. Moreover, Bayesian updating is compromised in the presence of wish fulfillment bias. When individuals are unsure of the relative value of various choices and must depend on loosely related cues as to what to choose, they can produce intransitive preferences, as Robert Goldstone (2015) suggests, along the lines of the Arrow paradox. Finally, regret theory, which has strong empirical support, exhibits intransitive preferences (Loomes and Sugden, 1982; Sugden, 1993). The rational actor model simply does not apply to such cases. In general, it must be determined empirically where the model applies and where it does not.

Should those concepts that are not amenable to mathematical formalization be obliged to reference an analytical core? Of course not. Some of the greatest sociologist made no use of any analytical models at all. Indeed, there are no mathematical sociologists among the great masters in the field. Perhaps there never will be. It is probably enough that one know how to use and interpret the mathematical models created by others. The reverse is true as well. Major analytical contributions to social theory are unlikely to come from theorists who lack deep intimacy with the workings of the phenomena with which they deal.

Let us take Alan Isaac’s example (Isaac, 2015) of the lifeworld concept, which was proposed by Edmund Husserl (1936/1970) for epistemological investigations, and further developed by Alfred Schutz (1932/1967) and Jürgen Habermas (1991) to study subjectively meaningful action, and is closely related to Pierre Bourdieu’s (1972/1977) notion of habitus. The lifeworld concept is an enriched version of the concept of belief in the rational actor model, with an associated notion that individuals who occupy similar social positions will tend to share similar representations of their natural and social world and their representations are derived from experience. This is a particularly welcome contrast to the common notion in sociological theory that beliefs are the product of the transmission of the cultural system to individual minds through the socialization process (Parsons, 1951). The lifeworld concept is also related to the theory of Bayesian updating in rational choice theory, because the basic idea is that the only access to new knowledge that we have is by transforming previous beliefs based on new experiences or mental rearrangements of existing beliefs.

I have benefitted greatly from exploring the works of prominent phenomenologically-oriented social theorists, and hopefully my own work has been enriched by their insights. A problem with their approach, though, is that much behavior is the product of mental processes that rarely or never reach the level of conscious perception and yet from the behavioral point of view, they nevertheless contribute to the system of beliefs of the individual actor. Often
such premordial beliefs become conscious only when they are violated. Decision theorists call these “zero-probability events,” to which Bayesian updating does not apply (Brandenburger, 2014).

4 Final Observations

Alternative models can be in conflict, researchers can be competitive in their search for scientific breakthroughs, research institutions can vie for research funds, and researchers can compete for jobs. But there are no scientific grounds for competition among academic disciplines. Such competition, which surely exists and is a profound impediment to advances in the behavioral sciences, is simply pathological.

There are in fact basic inconsistencies in the core principles that unite the common (or at least majority) beliefs of workers in different disciplines. I have outlined some of these inconsistencies (Gintis, 2007b) and devoted a whole volume (Gintis, 2009b) to addressing the problem as to what must change to bring contemporary economic theory in line with known facts and strongly supported theories from other behavioral disciplines, with special emphasis on sociology. I have similarly suggested how sociological models can be incorporated into population and evolutionary biology (Gintis, 2003; Bowles and Gintis, 2011), how results from evolutionary anthropology can be deployed to understand human political dynamics (Bowles and Gintis, 2005, 2007; Gintis and van Schaik, 2014), and how human sociobiology and inclusive fitness population biology can be integrated, creating a consistent unified theory (Gintis, 2014). It appears clear to me that the behavioral sciences will increasingly benefit from interdisciplinary borrowing of ideas, the possibility of which depends on vigorously attacking and resolving theoretical inconsistencies among disciplines. Treating our contribution in these pages as interdisciplinary conflict, however, does not advance this project.

Several commentators have framed Dirk Helbing and my suggested analytical core for sociology as a conflict between economics and physics on the one hand, and sociology on the other. Such terms as “disciplinary imperialism” (Witt, 2015) and “Trojan horse” (Hodgson, 2015; Macy, 2015) are indicative of this framing. Also indicative are critiques that assert that our suggestions will be rejected by sociologists because they run counter to the contemporary culture of the discipline. For instance, Michael Hechter (2015) writes:

[The Gintis-Helbing] formulation makes the reader wonder about the intended audience for the paper. Most sociologists are uncomfortable with the concept of equilibrium, and mathematically unprepared to understand the Walrasian model. So if the authors are aiming to reach a sociological audience, as their title seems to imply, this paper is virtually certain to fall on deaf ears.
I think it is extremely important to address the problem of how scientific advances, including in our case evolutionary game theory, complexity, agent-based modeling, and large-scale social dynamics, can be successfully introduced into a field the majority of whose members are not culturally or technically prepared for them. But it has been done many times in the past, and there is reason to believe that it will occur even more rapidly in the future. Catherine Eckel and Jane Sell (2015) have documented that much of the theory we offer is already represented within the field. Our belief is that once misinformation is cleared away (which is of course what our paper has tried to do), a firm basis for an analytical core for sociology will emerge. Why? Because the scientific method has been conducive to the emergence and acceptance of true theories in the past, and there is no reason that this will not continue to be so in the future.

A clear impediment to transdisciplinary work is the oft-expressed notion that there are certain concepts that are appropriate in one discipline but prima facie inappropriate in another. An example referred to in comments on our paper is that techniques from physics, as opposed to biology, are inappropriate in the behavioral sciences (Mirowski, 1991). Some commentators have followed this flawed either/or logic by asserting that there is no role for equilibrium theory in dealing with human society, which is a system of high biological complexity.

This either-or argument is clearly incorrect. The notion that physical theories are not applicable to biological entities is not accurate. Consider, for instance, the application of statistical physics to biological systems (Prigogine and Nicolis, 1977; England, 2013). Moreover, there is no conflict between the notion that societies are complex adaptive systems and that they can be modeled using dynamical systems theory in which equilibrium states exist. Models of species behavior in population biology and species interaction in ecology regularly use the notion of population equilibrium. Stephen Jay Gould and Niles Eldredge (1977) have suggested the term punctuated equilibrium to represent the phenomenon of species that change very little over very long time periods. For the general notion of the core genome of a species, a quintessentially equilibrium concept, see Gintis (2014). Finally, as Nowak, Andersen and Borkowski (2015) note, even individuals, the preeminent product of evolution, are often organisms with a stable equilibrium states described by homeostatic dynamics (Cannon, 1929).

References


Heyd, D. 2012. “Supererogation”.


