Department of Economics

Working Paper

The Sea Battle Tomorrow:

The Identity of Reflexive Economic Agents

By

John B. Davis

Working Paper 2020-01

College of Business Administration

Plenary paper presented at the Third International Economic Philosophy Conference, Aix-en-Provence, France, June 2016.

The Sea Battle Tomorrow:

The Identity of Reflexive Economic Agents

John B Davis

Marquette University and University of Amsterdam

John.davis@mu.edu

Abstract: This paper develops a conception of reflexive economic agents as an alternative to the standard utility conception, and explains individual identity in terms of how agents adjust to change in a self-organizing way, an idea developed from Herbert Simon. The paper distinguishes closed equilibrium and open process conceptions of the economy, and argues the former fails to explain time in a before-and-after sense in connection with Aristotle's sea battle problem. A causal model is developed to represent the process conception, and a structure-agency understanding of the adjustment behavior of reflexive economic agents is illustrated using Merton's self-fulfilling prophecy analysis. Simon's account of how adjustment behavior has stopping points is then shown to underlie how agents' identities are disrupted and then self-organized, and the identity analysis this involves is applied to the different identity models of Merton, Ross, Arthur, and Kirman. Finally, the self-organization idea is linked to the recent 'preference purification' debate in bounded rationality theory regarding the 'inner rational agent trapped in an outer psychological shell,' and it is argued that the behavior of self-organizing agents involves them taking positions toward their own individual identities.

Keywords: reflexivity, Simon, Aristotle identity, self-fulfilling prophecy

JEL codes: B11, B25, B41

1 The identity of reflexive economic agents

Reflexive economic agents continually revise their expectations of the future and continually adjust their behavior to those changing expectations. This paper develops a conception of reflexive economic agents as an alternative to the standard utility conception of economic agents, and explains their individual identities terms of how they adjust to change, rather than in terms of fixed preferences. The idea that identity might lie in change rather than fixity may at first appear paradoxical, but I draw on Herbert Simon and the idea of self-organization to motivate it. Simon dismissed the exogenous preferences utility function representation of agents as inadequate to the task of explaining behavior in complex, changing environments (Simon, 1956, 138), and recommended we replace it with a conception of endogenous agents whose "behavior is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1991, 7). I use this idea to extend Simon's idea of self-organizing systems to agents, reasoning that just as he thought we ought to explain the behavior of complex systems "in terms of the concepts of feedback and homeostasis" (Simon, 1962, p. 467), so we should also explain the nature of economic agents "in terms of the concepts of feedback and homeostasis." Indeed, I argue that treating complex economic processes as reflexive and self-organizing entails we also should explain agents and their behavior as reflexive and self-organizing.

Needless to say, the identity focus of this paper departs from the main concern with Simon's thinking associated with his idea of bounded rationality (cf. Grűne-Yanoff, Marchionni, and Moscati, 2014).[1] I agree that explaining bounded rationality is important to economics, but my view is that a bounded rationality and a bounded individuality are two sides of one issue (Davis, 2015). Since the utility function representation of individual identity is formally derived from the axiomatic representation of preferences, behavioral anomalies associated with the latter imply we lack an adequate definition of what individuals are. Indeed, how boundedly rational individuals can or should be explained is precisely the issue in a recent symposium and exchange in *Journal of Economic Methodology* between Gerardo Infante, Guilhem

[1] See the special December 2014 issue of the *Journal of Economic Methodology* on bounded rationality and the Grűne-Yanoff, Marchionni, and Moscati introduction to the issue.

Lecouteux, and Robert Sugden and Daniel Hausman regarding 'preference purification' – or how we might "reconstruct individuals' underlying or *latent* preferences by simulating what they would have chosen, had they not been subject to reasoning imperfections" (Infante, Lecouteux, and Sugden, 2016a, p. 6; cf. Hausman, 2016; cf. Infante, Lecouteux, and Sugden, 2016b).[2] The symposium debates whether bounded rationality implies a dualistic view of the self, and focuses on the idea that agents weigh their options rather than simply respond to them. In my view, this asks, in Simon's terms, whether boundedly rational individuals are self-organizing, and this is the issue I consequently investigate in this paper, if in a different connection than the symposium.

The argument of the paper, however, does not start with individual agents but with discussion of two characterizations of the economy as a whole – the standard static equilibrium conception and an alternative dynamic economic process conception. My view is that the problematic character of the utility conception and the advantages of a reflexive agent conception are respectively tied to the problematic character of the standard static equilibrium conception and the advantages of the dynamic process conception. I thus begin in section 2 with a critical review of standard equilibrium thinking using Aristotle's classic sea battle tomorrow problem, and argue that the standard view employs an equilibrium-shock model that cannot explain time in a before-and-after sense because it employs a closed systems view of the economy. In section 3, I turn to the idea of a reflexive economic process, and use the truth-reversing properties of self-fulfilling prophecies as a special kind of reflexive judgment to show how reflexive economic systems are open process systems in which behavior has a before-and-after character. I provide a causal model of how a reflexive economic system operates through feedback channels, characterize open systems as non-ergodic on reflexivity grounds, and finally return to Aristotle's sea battle problem.

Section 4 discusses the utility conception of the agent, and argues that if we say that the economy functions as an open, reflexive process, then it is misguided to think preferences should be complete, unless agents construct them as such themselves. Alternatively, it seems we should rather ask what sort of choice behavior is consistent with how we

[2] The expression 'preference purification' is Hausman's (2012), though see his many caveats, especially in Hausman (2016).

understand action and time. I characterize this behavior as adjustment behavior, which introduces the following section's treatment of reflexive economic agents. In section 5, following Simon, I first explain adjustment behavior in terms of its moving to stopping points. I then advance a general account of individual agent identity in which an endogenous shock event disrupts an existing basis on which agents' individual identities operate, and their adjustment involves them self-organizing themselves on some new basis on which their individual identities subsequently operate. I give four examples from the literature to show how different models of behavior emphasize change in the sub-personal and/or supra-personal dimensions of identity, and then use the idea of self-defeating prophecies to characterize agents' own orientation on their identities. Section 6 comments briefly on the paper's arguments.

2 Standard theory's equilibrium-shock model and time

The standard Nash definition of equilibrium in economics is defined as a state of affairs fully at 'rest,' meaning no agent has an interest in deviating from the allocation of resources and strategies associated with that state. That is, it is a state of perfect coordination of all agents' plans and the idea of a perfectly static state of affairs. Consider the application of this conception to the Walrasian understanding of a general equilibrium of markets, the dominant framework employed by economists to explain the market economy. Equilibrium is then a perfectly static state of affairs in the infrequently appreciated sense that, according to the Sonnenschein-Mantel-Debreu results regarding multi-market general equilibria, equilibria generally cannot be shown to be stable (Rizvi, 2006). This means that the standard equilibrium conception cannot explain movements from out-of-equilibrium to equilibrium, or how an economy gets into equilibrium, and thus refers to a perfectly static state of affairs in the further sense that it lacks any *internal* principle of motion. An equilibrium just is, full stop, as shown by the fact that only existence (and even not the uniqueness) of equilibrium can be shown. This renders comparative static analysis, the work-horse of standard theory, essentially meaningless, because comparative static analysis is about getting into a new equilibrium given a 'shock' to an old equilibrium. But if the theory lacks any *internal* principle of motion associated

with how an economy gets into equilibrium, does the idea of shocks suggests a theory of *externally* caused motion associated with how an economy can get *out* of equilibrium?

My hypothesis is that the standard view of equilibrium as a perfectly static state of affairs with no internal principle of motion renders the 'equilibrium-shock' model of external motion philosophically problematic. To argue this, I claim that the 'equilibrium-shock' model of motion cannot address an ancient philosophical problem associated with the relation of truth claims to time, which first emerged as the problem of future contingents advanced by Aristotle in his famous sea battle tomorrow example (Aristotle 1963). Future contingents are statements about future states of affairs that are neither necessarily true nor false today. For Aristotle, that a sea battle will not be fought tomorrow is neither necessarily true nor false today. Suppose, then, that a sea battle will not be fought tomorrow. If a sea battle will not be fought tomorrow, then it was true yesterday that it will not be fought tomorrow. Yet since all past truths are necessary truths, it must also be necessarily true today that a sea battle will not be fought tomorrow. However, this conclusion is fatalistic and runs counter to our intuitions about the future being open. Thus any theory that employs future contingents needs an answer to Aristotle.

Consider the standard 'equilibrium-shock' model of external motion regarding how an economy can get out of equilibrium. A shock is an event in time because it differentiates before and after. On the one hand, then, from the perspective of a given equilibrium, a shock event is a future contingent state of affairs, something that could occur, and as such its occurrence should be neither necessarily true nor false. On the other hand, given that an equilibrium is a perfectly static state of affairs, shocks are fully external to any given equilibrium configuration. Thus from the perspective of any given equilibrium configuration, shocks necessarily do not occur. Equilibrium is forever. But then without shocks there is no differentiation of time into before and after, so the 'equilibrium-shock' model fails as a theory of externally caused motion. Note also that the failure of this conception is due to the lack of any internal principle of motion in the standard equilibrium conception, which as a fully complete, static representation of the economy, closes off any role for external causal factors. Just as necessarily there can be no sea battle tomorrow, so there can never be equilibrium shocks tomorrow.

Aristotle, in fact, similarly diagnosed the problem of future contingents as a problem of completeness, specifically, completeness with respect to the scope of application of the logical principle of bivalence – the idea that every statement must be either true or false – to any and all statements irrespective of their temporal dimensions. His solution to the problem and escape from fatalism was to say that the principle of bivalence applies to the future differently than it applies to the past and the present, though what he meant by this and what philosophers have argued this could mean is much disputed in the history of logic and philosophy (see Rice 2014), and will not detain me here. Instead, in the next section I will discuss why we cannot always say that a statement about the future is true or false when we operate with an open process conception of the economy, and here only comment on why it may seem odd for me to have used the problem of future contingents to comment on standard equilibrium theory.

That oddness, I believe, comes from combining a discussion of how truth is determined with the mathematics of equilibrium determination. A Nash representation of a set of equilibrium strategies models a mathematical solution to a set of behavioral functions. That it 'models' that solution and its attendant representation of 'behavior' tells us that the empirical truth or falsity of the propositions involved is of no particular importance, and indeed can even be set aside. Rather, the main thing that is important about that representation of agents' behavior is that it be mathematically consistent in the sense of producing a solution to a set of equations.[3] Indeed, the mathematical utility function representation of agent behavior is not meant to be evaluated according to how well is describes people's behavior, but according to its mathematical tractability, so the common complaint that this representation of agents is unrealistic basically aims at the wrong target, at least according to its proponents.

The problem that Aristotle's sea battle problem points us towards, then, is the sharp divorce in mainstream economics between the logic of truth and the mathematics of equilibrium determination. What I conclude from this, however, is not that we ought to abandon mathematical representations, nor certainly that claims about the realism of

[3] Hausman makes essentially this same point in connection with his distinction between theories and models (Hausman, 1992, 70-82). Boumans does as well in discussion of models (Boumans, 2015). Lawson makes the point in relation to the goals of consistency and realisticness (Lawson, 2013).

economic theory are of no philosophical importance, but rather that realism and a concern with how truth is determined ought to constrain and determine the nature of mathematical reasoning in economics. In particular, then, I recommend that we abandon mathematical representations of the economy that preclude employing before-and-after treatments of time in favor of mathematical representations that allow for this.[4]

However, I leave the task of advancing alternative mathematical representations of the economy to others, and in the following section give a philosophical characterization of the economy as an open or endogenous process rather than as the closed system such as standard equilibrium theory employs. There of course exists considerable philosophical and methodological literature regarding the distinction between open and closed types of systems in economics, much of which emphasizes uncertainty, but I adopt a somewhat different entry point on the subject from many others by developing the open-closed distinction in terms of the idea of reflexive economic processes driven by the behavior of reflexive economic agents. This will in turn introduce my discussion of reflexive economic agents in section 5. In the following section, then, I will emphasize how an open reflexive economic process conception makes action and time central to the explanation of economic systems, and offers one way of addressing Aristotle's problem of future contingents.

3 The reflexive economic process conception and time

A reflexive economic process is one in which agents form expectations and beliefs about the future based on their understanding of the world, and this influences their actions in the present, which in turn influences future states of the world. The conception of the world as a reflexive economics process is thus a conception of action framed in before-and-after terms. All economics, then, is in principle concerned explaining the world as a reflexive economic process, since economic agents are assumed to form expectations and beliefs about the future that affect their actions in the present.

[4] In my opinion (and it is just an opinion), an alternative mathematical representation of the economic process that accommodates a before-and-after treatment of time involves an algorithmic type of mathematics (cf. e.g., Velupillai, 2011).

At the same time, the standard rationality theory treatment of equilibrium as a state of affairs ultimately at 'rest' negates this before-and-after temporal dimension of action, both in macroeconomics via the idea of rational expectations and in microeconomics via the idea of optimal or Bayesian (and least-square) learning. In the macro case of rational expectations agents' expectations regarding the future are on average consistent with the correct model of the world. In the micro case of optimal or Bayesian learning agents' rational beliefs about the future smoothly converge on the correct model of the world. When agents' expectations and beliefs regarding the future are rational or on average consistent with the correct model of the world, and when agents' converge smoothly on the correct model of the world, then the economy simply achieves the equilibrium values inherent in agents' model of the world as if time did not matter. Nominally agent's expectations of the future still influence their actions in the present, but this occurs in such a benign way that one can ignore it and proceed as if there were no temporal dimension to behavior.

This standard view, then, can nonetheless be represented in causal terms so as to distinguish the direct effects of people's actions on the world associated with agents' models of the world and a feedback channel associated with how agents' expectations/learning affect their models of the world. Agents' actions *a* can then be said to have direct effects on the world *b*, or *a* -> *b*:

$$a \rightarrow b \qquad [1]$$

Thus, [1] represents agents' model of the world. When agents form rational expectations or optimally learn about this causal relation *a* -> *b*, and act on this basis, then the *a* -> *b* relation acts reflexively on itself, and makes that model of the world self-confirming:

$$a \rightarrow b \rightarrow (a \rightarrow b) \qquad [2]$$

Then the combined overall effects, (=>), of the direct causal model [1] and the reflexive feedback channel [2] produce both *b* and (*a* -> *b*):

$$a \text{ and } a \text{ -> } b \text{ and } a \text{ -> } b \text{ -> } (a \text{ -> } b) \text{ => } b \text{ and } (a \text{ -> } b) \qquad [3]$$

Consequently, since *a* -> *b* and (*a* -> *b*) exhibit the same direct effect of *a* on *b*, the feedback channel only plays a nominal role that can be ignored, the process is closed, and the passage of time is effectively negated.

Contrast this with the case of an open, endogenous economic process where expectations are not rational and learning is not optimal. Agents' actions *a* have direct effects on *b*, but since agents' expectations and learning do not confirm [1], the feedback channel changes the nature of the relation between *a* and *b* such that [2] is replaced as follows:

$$a \text{ -> } b \text{ -> } (a \text{ -> } b)' \qquad [4]$$

Replacing the *a* -> *b* relation by the (*a* -> *b*)' relation, [3] is then replaced as follows:

$$a \text{ and } a \text{ -> } b \text{ and } a \text{ -> } b \text{ -> } (a \text{ -> } b)' \text{ => } b \text{ and } (a \text{ -> } b)' \qquad [5]$$

In this case, time operates in a substantial, before-and-after way on account of the changed feedback channel. Were [5] to be the general case, and [3] a limiting case, then across a sequence of periods in time agents would need to constantly adjust their causal

models of the economy: (*a* -> *b*)', (*a* -> *b*)", etc.[5]

The world in this second case, then, is open in the sense of being nonergodic and path-dependent due to how reflexive agents' changing expectations and consequent actions alter the relation between *a* and *b* in time.[6] The limiting case rational expectations/optimal learning view produces a closed systems approach because the allowed operation of reflexivity only confirms the existing model of the economy. It also allows the *a* -> *b* causal model of the economy to be reduced to a type of mathematical expression which omits any real incorporation of time. All other types of expectations and non-optimal learning are open systems approaches in that the operation of reflexivity requires dynamic representations of the economy in before-and-after time via how agents' expectations and models of the economy are continually revised in terms of one another. The principle of motion that these dynamic pathways exhibit is an endogenous one in that the operation of reflexivity ensures that what occurs at one point in time influences what occurs at a later point in time. Time is not an unconnected sequence of independent states, as it must be represented in the equilibrium-shock model according to the mathematical treatment of 'shock' as an 'event' outside the model, but rather a connected process appropriate to our before-and-after representation of time. What is principally different, then, about a reflexivity-based treatment of the open-closed systems distinction is that it makes action and time central. That is, a reflexivity-based treatment of the open-closed distinction is both an ontological treatment of that distinction, because of the role of action, and also an epistemological one, because action is predicated on agents' state of knowledge.

[5] Simon's two blades process analysis is richer than [5] since he allows for the case in which a change in the environment occurs independently of the effects of the feedback channel. I leave this additional layer of complexity aside in order to emphasize the role of the feedback channel in order to focus on the behavior of reflexive agents in section 5.

[6] In Paul Davidson's terms, the world is 'transmutable' (Davidson, 1996). At the same time, the world is only 'evolutionary' in the loosest sense of the term. Evolutionary processes, at least in the classical Darwinian sense, basically occur 'behind the backs' of agents, whereas reflexive systems depend on how agents' actions influences the economy's time path. See Barkley Rosser's careful discussion of the meaning and interpretation of the concept of nonergodicity in Post Keynesian economics emphasizing its relation to the concepts of non-stationarity and non-homogeneity (cf. Rosser, 2015).

But this causal analysis leaves unexplained just how 'open' an economic process understood as endogenous and reflexive can be. Action influences the world and its time path. Yet it could still be the case that the economy is an endogenous process due to less than rational expectations, is a nonergodic system because action causes the phenomena to be less than stationary, and yet the effects of agents' behavior through the feedback channel in the expectations-models adjustment process are sufficiently modest that it still largely appears as if the world is an ergodic system in which time does not matter. After all, as we know, in a large number of respects the world works pretty much the same way day-in-day-out.

This then calls for closer attention to the nature of the expectations feedback channel, and to do this I emphasize its agency-social structure character (Archer, 1995; Lawson, 1997), and illustrate this with a special case, the now classic, highly stylized example of how a feedback channel has significant effects on an economy's time path, namely, Robert Merton's (1948) treatment of a bank run as a self-fulfilling prophecy (SFP). In his famous Depression bank run example, a bank examiner mistakenly judges a bank will become insolvent (an endogenous 'shock'[7]), people act in conformity with this judgment causing a bank run, and the bank becomes insolvent, thus fulfilling the examiner's prediction or prophecy. Rather, then, than explain the feedback channel in a purely epistemological way in terms of only changes in beliefs or expectations, Merton characterizes the situation in agency-social structural terms involving the agency interaction relationship between the bank examiner's actions and depositors' actions which is embedded in and acts on the social-institutional structure that determines how the banking system works. All this, then, is what underlies the bank causal model *a* -> *b* that agents work with and the expectations feedback channel *a* -> *b* -> (*a* -> *b*)' that agents exercise when that causal model is called into question.

If, then, a causal analysis of the economy as an open, endogenous process leaves unexplained how 'open' an economic process might be, what a SFP does is provide a clear measure of openness in the form of a truth reversal of agents' judgments, where this reversal in turn is the product of a whole set of changes in the attendant agency-social

[7] It is endogenous because it arises from a judgment internal to the system as compared to an exogenous shock brought about by an event outside that system.

structure relationships. The bank examiner mistakenly judges the bank to be insolvent when it is solvent, and this causes the bank to become insolvent on account of the nature of the interaction between the examiner and depositors in the context of how the banking system works. What was false is taken by depositors to be true, and then after the bank run it is indeed true that the bank is insolvent because this agency-structure interaction has changed what is true.

Of course the Merton bank run example is highly simplified case that exaggerates how clearly agents' interaction affects what is true. More often than not changes in what is taken to be are not clear since our views about what is the case in the world with complex social structures involves many claims and assumptions whose interconnected nature makes it difficult to evaluate individual truth claims. But it would be a mistake to infer that this Merton's extreme truth reversal case is therefore unlikely to occur or that when it does it is only on a small scale. As everyone now knows, the recent financial crisis was essentially a banking crisis quite parallel to Merton's bank run example in which banks were judged solvent until they were successfully shorted and indeed became insolvent. Those who shorted the banks, that is, played the role of Merton's bank examiner,[8] and the crash in the overnight lending market played the role of Merton's depositors.

So openness should be characterized not only in truth-functional terms but also in terms of the social stability of beliefs. Depending on the domain, people's beliefs are more or less secure, and thus more or less subject to, or vulnerable to, a revision process in which agents search for causal models *a* -> *b* and investigate possible feedback channels *a* -> *b* -> (*a* -> *b*)' regarding their strategies for revising those models. This makes just how 'open' an economic process is depend on the way in which the agency-structure relationships are embedded in social institutions – obviously still a quite 'open' matter. But this, I suggest, is what one should expect when one takes the economy to be an open, endogenous and path-dependent process.[9]

[8] The nature of a 'short' is to bet against the consensus view, or what is taken to be true, and a successful short changes that view.

[9] How much lock-in, then, economic processes exhibit (David, 1985) would seem to depend in part on how durable the social-institutional foundations of social interaction are.

This analysis, then, also points to one advantage of an open systems, agency-structure approach to the economy over a mathematical representation of the economy. A mathematical representation of the economy as a complex dynamic process can map changes in variables, and explain how an economy works in unexpected ways in terms of the idea of phase transitions. But when do mathematically-described phase transitions count as the economy working in an unexpected way and when do they count as it working in pretty much the same way? In thermodynamics, a phase transition involves a clear qualitative change since it involves a change from one state of matter to another, which clearly marks out a before-and-after sequence. What counts as a change from one 'state of matter' to another in the economic world? In social science changes in 'states of matter' are subject to judgments regarding what differences those changes make to agents. Thus agents' judgments about what is true or false ultimately determine what counts as a phase transition, as when we say the characteristics of the economy after the financial crisis are not true of the characteristics of an economy before the crisis.

From this vantage point, it seems a rather straightforward interpretation of Aristotle's sea battle tomorrow problem is as follows. Whether there is a sea battle tomorrow affects many people, and thus many people would form expectations about this possibility. Agents' expectations of possible sea battle tomorrow, then, are likely to influence other agents' actions. Should these actions fulfill or defeat the expectation of a sea battle tomorrow would then determine whether a sea battle occurs tomorrow. Thus the fatalism paradox that Aristotle suggested fails since whether a sea battle occurs tomorrow is not inevitable but depends on the actions people undertake. Indeed, Aristotle likely posed the fatalist view as a *reduction ad absurdum* argument to show that the claim a sea battle would or would not occur as inevitable was false and the paradox is not a paradox. Accordingly, the key assumption the paradox makes that Aristotle likely believed to be false was that action and truth are independent. If they are not, then action can change what is true, and the future is open, not inevitable.

4 The utility conception of the economic agent and the completeness assumption

I have claimed that rational choice theory's standard utility function conception of the economic agent goes hand-in-hand with standard theory's equilibrium-shock model of the economy. Rational choice theory is a normative theory of choice in that the utility conception of the agent depends on axiomatic foundations which discipline what choices agents ought to make. The axioms were originally taken to be self-evidently true, but the consistency in choice they produce is as much a matter of consistency with the equilibrium-efficiency properties of the standard equilibrium theory. The purpose of the axioms governing utility functions, that is, is not just to prescribe consistent choices, but to prescribe equilibrium-efficient consistent choices. But if standard equilibrium theory is questionable in regard to our understanding of action and time, then searching for axiomatic foundations for utility functions and rational choice seems misguided – the project aptly dubbed the 'preference purification approach.'[10] Rather, one ought to ask what behavior is consistent with how we understand action and time, and then explain the foundations for that behavior with an appropriate understanding of the agent.

Consider the important completeness axiom, which says that the agent must be able to compare any two imaginable states of the world, x and y, by a relation of preference R. If the axiom does not hold, and individuals' preferences are incomplete, much of standard theory is called into question, including welfare analysis and the WARP axiom of revealed preference theory. Why then might it fail? The most widely held answer is vagueness, or that agents may be unable to clearly determine their preferences regarding vaguely specified alternatives (Broome, 2004). How, then, ought researchers understand vagueness? The dominant response seems to be to investigate how though preferences might be vague, but the completeness axiom can still be retained, or at least some correlate assumption about preferences holds that largely serves the same purpose so that agents' choices are still effectively rational. But I believe this strategy gets things backward. Rather it would seem to make more sense to address incompleteness straight on, and explain decision-making behavior on that basis, as does much of behavioral economics, which has unmoored itself from rational choice.

[10] See footnote 2.

The motivations of behavioral economists, however, differ from mine. They largely set aside the issue of how the economy as a whole should be represented in the interest of achieving greater ground-up realism regarding choice. In contrast, if how we explain behavior depends on how we explain the economy as a whole – in equilibrium terms or process terms – then we ought to investigate vagueness and incompleteness as reflective of agent behavior when we operate with a process conception of the economy. This is indeed what the causal model analysis of the last section implies. Since the feedback channel,

$$a \text{ -> } b \text{ -> } (a \text{ -> } b)' \qquad [4]$$

alters the *a* -> *b* relation, there is good reason to suppose that the *b* terms in *a* -> *b* and (*a* -> *b*)' are not comparable or perhaps only vaguely comparable. If the world has not changed too much, then the completeness axiom might be retained not as an axiom but as an observed relationship. Its basis, that is, would not be the requirements of equilibrium theory, but an assessment of how open or closed the world is in particular circumstances when we explain the economy as an open reflexive process. That assessment would also entail giving attention to the agency-structure background of the change in question, since how inertial change is, and how complete preferences might be, depends on how robust institutions governing social interaction are in making tomorrow more or less like today. So in contrast to behavioral economists, I give the issue of completeness reflexivity foundations, not psychological ones.

However, the general caution regarding the nature of preferences that behavioral economists have advanced, that preferences are often not stable for a whole variety of reasons, is still worth attention. Menu dependence, for example, tells us that a given set of preferences need not be stable when events occur that alter menus. However, in my view the more difficult issue here concerns preference formation. If the world is non-ergodic, then new states of the world regularly appear. To suppose that options in new states of the world can always be compared based on sets of preferences used to compare

options in past states of the world is heroic at best, since completeness has to then be defined as an unknown ability individuals have that is everywhere and always versatile and competent. This mystery flies in the face of clear, every day evidence that some market participants actively work to manipulate other market participants' preferences (by often exploiting menu dependence). Of course the idea that preferences are not just formed but are actually constructed by social forces has been around for a long time, and indeed has also been off the agenda of mainstream economics for a long time. Why is this when the world outside of economics does not regard preference manipulation as controversial? Why do most economists neglect it? One answer is that questioning preference autonomy undermines the independence of supply and demand and the supply-and-demand balance on which standard economics relies. But that balance is premised on the equilibrium-shock model being a correct since it is that balance which mathematically 'closes' the model. Doubts on this score consequently unravel all that depends on it right back to how we should interpret the completeness assumption.[11]

An alternative view of completeness is John Searle's view that complete preferences are "the product of practical reasoning" and not given characteristics of the individual (Searle, 2001, 253). I comment further on the identity implications of this idea below, and here only connect it to the idea of a reflexive economic process. If we regard a reflexive economic process as one that continually changes the world, then essentially preferences are continually being made incomplete, so that the issue is rather how and whether they might become complete through the actions of agents. This then makes adjustment behavior central to our characterization of economic agents, and consequently I now go on to how such agents can be understood both in terms of adjustment behavior and in terms of identity.

5 The behavior and identity of reflexive economic agents

[11] Questioning preference autonomy, clearly, also jeopardizes the basis on which individual identity operates in the utility conception – in a circular way I have argued (e.g., Davis, 2011, pp. 6ff). If individuals' sets of preferences are what constitute their individual identities, and their preferences are influenced by others, what kind of 'individuals' are they? The problematic 'inner self' view is one strategy to respond to this.

I argued at the outset that explaining complex economic processes as reflexive and self-organizing entails explaining agents in such processes as reflexive and self-organizing. My goal in this section of the paper, then, is link my section 3 account of reflexive economic processes to an explanation of the behavior and identity of reflexive economic agents in terms of their ability to adjust the grounds for their behavior in a self-organizing way in response to change in their environments. I see two steps involved to doing this. First, I need to explain agents' adjustment behavior in terms of how it continues to a stopping point. Adjustment processes are not open-ended, as Simon made clear with his satisficing idea. They involve a response to an event that initiates them, and they run their course when agents have adjusted to that event. Second, I need to explain how agents' adjustment behavior causes them to self-organize their identities as individual economic agents at such stopping points. To explain the idea of agents' identities as self-organizing, I argue that reflexive processes can disrupt agents' individual identities, and their adjustment and self-organization comes about through their re-organization of these identities. I illustrate this first using Merton's bank run example, and then generalize that explanation to three additional examples.

First, to explain adjustment behavior, I follow Simon's explanation in terms of satisficing, which he understood to be a matter of reaching an aspiration level. The two questions Simon's explanation naturally raises are: how are aspiration levels set, and how does one know when they are achieved? His 'two blades of the scissors' answer was that agents' aspiration levels depend on the type of process in which the agent is involved, and their achievement depends on how the agent adjusts within that process (Simon, 1955, 111-113).[12] Consider, then, the reflexive economic process that I outlined above in connection with Merton's SFP bank run example. There the bank examiner's evaluation of the bank affects depositors' view of the bank's solvency, this feeds back on and changes their causal model of the bank and also their behavior, and this produces the bank run. Depositors' adjustment behavior is thus caused by the reversal in judgment about the bank's solvency. This then tells us both how aspiration levels are set and how one knows when they are achieved. The new judgment that the bank is insolvent when it was previously believed

[12] In terms of process, Simon argued that decision makers' aspiration levels rise or fall as the alternatives they face are respectively easy or difficult to discover.

solvent sets depositors' new aspiration levels (to withdraw all their deposits) and also drives their adjustment behavior (withdrawing their deposits) to the stopping point (full withdrawal of deposits) at which they achieve their aspiration level.

We can increase the realism of this account by explaining how this reflexive process works in agency-structure terms. At the outset, the bank examiner and depositors interact in a socially structured way determined according to how banking laws work, monetary systems, and financial markets are organized. The basis on which they interact is then disrupted by the agency of the bank examiner (agency affecting social structure). The ensuing withdrawal adjustment process on the part of depositors reflects the effects of their re-aligning their judgments about the bank to the bank examiner's judgment. In effect, then, the bank examiner sets an aspiration level inconsistent with the existing basis on which the examiner-depositor relationship operates, so depositors' withdrawals must proceed until the results of their actions conform to this new aspiration level, and thereby establish the new basis on which the examiner-depositor relationship operates. At the same time, the banking system and how markets are organized has been affected by the bank's failure. In agency-structure terms, the agency of the bank examiner has influenced social structure, and the consequent adjustment within that structure has changed the basis on which this agency-structure relationship will subsequently operate.

Second, then, to explain how agents' adjustment behavior leads them to self-organize themselves as individual economic agents, note how in Merton's example the bank examiner's evaluation affects depositors' identities. Before the evaluation becomes known depositors' identities are tied to their individual interests alone, since the earnings on their deposits accrue to them individually. However, when that evaluation becomes known, depositors recognize they then also have social identities as depositors since each has the same interest as every other depositor, and each acts on the same basis as every other depositor in withdrawing their funds. That is, all depositors now see themselves as representative agents of the group of depositors, and act as a representative depositor would act. Yet at the same time, though depositors adopt this depositor social identity, they nonetheless still retain and are still motivated by their individual identities, since they know that if they fail to withdraw their own funds when others are withdrawing theirs, that they will lose their funds individually in the bank failure, and put their

individual identities at risk. They thus act on both identities, both in social identity terms as representative agent depositors and in individual identity terms as independent agents. However, when the bank run has run its course, their social identity as depositors ceases to be relevant, their individual self-interested identities fully occupy them. That is, they have self-organized themselves in terms of those individual identities. In Simon's terms, then, the stopping point in his satisficing-aspiration level explanation of an adjustment process is also a stopping point in agents' self-organizing identity adjustment.

The general view, then, is that an endogenous shock event disrupts an existing basis on which agents' individual identities operate, and their adjustment involves them self-organizing themselves on some new basis on which their individual identities subsequently operate. Below, I further explain the two sides of this analysis in reflexivity terms, the disruption side and the self-organizing side, by demonstrating the complementarity between SFPs on the disruption side and self-defeating prophecies (SDPs) on the self-organizing side. But to better prepare the ground for this I first give three more examples to generalize from the Merton example.

Consider, then, Don Ross's neurocellular account of individual identity. Ross argues that individuals are made up of collections of sub-personal neural agents or neurons, their sub-personal multiple selves, which interact in coordination games internal to the individual to produce individual identity (Ross, 2005, 2006, 2007). These sub-personal neural agents each act in their own interest, and compete for bodily resources as relatively independent agents and individual identities. The human body, of course, is regularly affected by any number of psycho-physical events that influence how these sub-personal agents interact and coordinate in order to serve their own individual identity interests. I characterize these events as endogenous shocks that disrupt an existing neural agent coordination, and the adjustment to a new neural agent coordination as the self-organization of a new individual identity. Ross's explanation is more sophisticated than this outline because he also discusses at length how the interaction between individuals compels each individual's sub-personal neural agents' coordination.[13] Nonetheless, his model illustrates the general one suggested above: an endogenous shock to an existing

[13] For example, see his discussion of language (Ross, 2007).

identity basis followed by self-organizing adjustment to a new identity basis. The main difference from the Merton example is in the forms of identity involved. If Merton's process goes from individual identity to individual identity with social identity to individual identity,

$$I \rightarrow I/SD \rightarrow I'$$

Ross's process,

$$I/MS \rightarrow I'/MS'$$

goes from one individual identity with multiple selves to a differently organized individual identity with multiple selves.

For a related individual identity/multiple selves case, consider the Santa Fe artificial stock market model developed by complexity researchers under Brian Arthur's leadership (Arthur, 1995; Arthur *et al.*, 1997). Profit-maximizing agents have the task of investing in a single asset, and form multiple subjective expectation models regarding what moves the market price of that asset. Different agents prioritize different models, and in a trading day the price of that asset moves to a value that reflects the distribution of these different expectation models across agents. How well agents do in terms of profits with their respective prioritizing at each stage of the process, then causes them to re-order their expectation models to improve performance in subsequent rounds. Seen in evolutionary terms, agents' models take on a life of their own as they effectively compete with one another through the intermediary of investors. These models effectively 'inhabit' investors, and are thus like Ross's sub-personal agents or neural multiple selves. One difference is that investors are ordered collections of such models whereas the ordering principle for Ross's whole individuals is a coordination equilibrium. In identity terms, however, the analysis is formally parallel – I/MS -> I'/MS' – though rules or expectation

models are not internal to investors in the way neural selves are internal to people. The Santa Fe artificial stock market model of course does not work as a coordination game, but rather as an evolving complex dynamic system. Its periodization through successive trading rounds nonetheless lends it a disruption-adjustment pattern that works through identity change in self-organizing investors' constant re-ordering of their expectation models.[14]

Last, moving away from multiple selves approaches and back to social identity approaches like Merton's, consider the Marseille fish market analysis developed by Alan Kirman and his colleagues (Kirman, 2011, 72-109). In the late 1980s the Marseille market was reorganized in such a way as to allow it to function as an open, arms-length auction type process as in standard competitive theory. What was observed, however, is that buyers and sellers instead formed close contacts, interacted directly rather than indirectly with one another, and developed preferences regarding partnering with some people versus others. That is, rather than a typical competitive auction process, trading networks emerged which segregated buyers and sellers into different loyalty relationships. These loyalty relationships, then, are a partnership type of social identity (rather than social group type social identity as in Merton's example) that individual buyers and sellers adopt alongside their individual identities. Thus, since buyers and sellers came to the market as individual identities under the disruption associated with the market's reorganization as an auction, their adjustment to the new market organization involves them layering these new loyalty social identities onto their individual identities,

I -> I/SD

and self-organizing themselves on that basis, unlike in Merton's example where they move back to individual identities alone. Later, this analysis was extended in a different

[14] More fully, since the market continually evolves, the expectation models that individuals are made up of also continually evolve. Thus the analysis allows that new expectation models may emerge as combinations of old ones, and the identity scheme is better represented as I/MS -> I'/MS' -> ….

context to the social group type of social identity, such as individuals joining groups that function like clubs (Horst, Kirman, and Teschl, 2007).[15]

What is the general form of the explanation, then, across these four examples? Turning to the two sides of the analysis, the disruption side and the self-organizing side, I argue that as reflexive processes they exhibit a complementarity which can be explained in terms of the complementarity between SFPs on the disruption side and self-defeating prophecies (SDPs) on the self-organizing side. A SFP, as in Merton's example, makes something false become true due to how people react to some event, whereas a SDP makes something true become false due to how people react to some event. The now classic stylized example in this case is the failure of the Y2K prophecy that computers would all fail at the beginning of the year 2000. It seems that this would indeed have happened, but the prediction that it would led computer engineers to re-appraise their existing *a* -> *b* model of computers, and act in such a way as to instantiate a new (*a* -> *b*)' model that secured computer systems against breakdown. Rather than the disruption that SFPs involve, SDPs thus deliver systems from disruption by making the undesirable, imminent, true states of affairs they involve false states of affairs through the actions agents undertake on their view of those undesirable state of affairs. In agency-structure terms, SDPs exhibit the tendency of social structures to re-stabilize themselves in response to endogenous agency shocks.

In identity terms, SDPs are an adjustment response to an identity disruption, whatever its origin, that precludes what would be the case from becoming the case. For Merton, depositors see their individual identities as linked to their social identities, and act so as to avoid losing their funds. For Ross, the appearance of new coordination equilibria for an individual's neural selves prevents the individual from being incapacitated. For Arthur and his colleagues, investors re-order their expectation models of the asset price to prevent themselves from losing money. For Kirman and his colleagues, buyers and sellers form loyalty relationships to forestall undesirable atomistic trading outcomes.

[15] Clubs exhibit high excludability and low rivalry. Horst et al. argue that adding new members changes the character of the club. Thus the social identity character of that club evolves as new individuals join it, just as the identity of individuals change when they become club members.

Seen in this way, however, there is an asymmetry between SFPs and SDP behaviors. Adjustment in the case of a SFP is reactive and backward-looking whereas adjustment in the case of a SDP is prospective and forward-looking. At the same time, the asymmetry between SFP and SDP behaviors goes beyond different orientations toward time because SDP behavior also involves individuals taking a position in regard their own identities framed by the goal of self-organization. This is where the debate in the *Journal of Economic Methodology* between Infante, Lecouteux, and Sugden and Hausman in my view comes into play. They ask whether there is some kind of option weighing activity in which individuals engage, the effect of which would be to remove the possibility that individuals are dual selves. Implicitly, then, they ask whether individuals can effectively 'step outside of themselves' to take positions on their own identities framed by the goal of self-organization.

What it consequently seems they are addressing is the need for some sort of hierarchical or multi-level conception of individuals with a reflexive capacity to manage their behavior based on the goal of individual self-organization. In SFP behavior, individuals' organization of their sub-personal and/or supra-personal identities is disrupted, but in SDP behavior individuals take a position toward this disruption of themselves, where that position involves setting out a behavioral course of action meant to re-organize their sub-personal and/or supra-personal identities.

My argument in terms of SFPs and SDPs, then, may seem unduly labored. However, its ultimate rationale is simply to argue that individual behavior and identity needs to be understood in terms of some sort of capacity to reflexively orient on that behavior and identity, a type of idea which has had little place in the theory of decision-making in economics, with a few exceptions. Most notable in this latter regard is Amartya Sen's representation of agents' overall identities that distinguishes on one level three different types of 'privateness' or kinds of individual identities that need to be managed (self-centered welfare, self-welfare goal, and self-goal choice) and on another level altogether a "fourth aspect of self" associated with being able to engage in "reasoning and self-scrutiny" and "examine of one's values and objectives and choose in the light of those values and objectives" associated with those types of privateness (Sen, 2002, 33-36; cf.

Davis, 2007, 2011).[16] Searle, I suggested, reasons in a related in way in his emphasis on practical reasoning, arguing that it is a mistake to think that complete "preferences are given *prior* to practical reasoning" when they are instead "the product of practical reasoning" (Searle, 2001, 253). In relation to individual identity, practical reasoning then involves taking a stance on one's identity in a self-organizing way.

My argument here, however, avoids claims about practical reasoning capacities, because it is framed only in terms of what we need to say about agents if we suppose the economy is a reflexive process. What I claim we need to say is that their adjustment behavior is self-organizing, and this involves not only their revising their behavior in light of their expectations of the future, but also revising it in such a way as to function as individuals.

6 A sea battle tomorrow?

My goal in this paper was to explain how the behavior and identity of individual reflexive economic agents might be understood as an alternative to the standard utility conception of agents. My explanation depends on understanding the economic process as reflexive, and I explain reflexive economic processes in terms of the adjustment behavior of reflexive economic agents. What I say about this process and agent conception is that they are both open in the sense that action affects the world, and thus occur in before-and-after time, so that what is the case in the world depends on action in a non-fatalistic way. So my assumption – one it seems not held by many economists – is that explanations in economics need to be adequate to our most basic views about the relationship between time and action. An implication of these views is that the future is not determined. Whether, as Aristotle asked, there will be the equivalent of a sea battle tomorrow depends on the actions people undertake in advance of tomorrow. This can be seen in how our judgments about what is true or false are susceptible to reversal according to what we do, as demonstrated, albeit in a highly stylized and simplified way, by SFPs and SDPs. Of course, many of our beliefs about what is true and false about the world are

[16] Self-centered welfare concerns one's own self-interest, self-welfare goal concerns one's own welfare, which may include own-welfare enhancing sympathy for others, and self-goal choice concerns one's own non-welfare goals. See Hedoin (2016, forthcoming) for a discussion of the relation of Sen's different levels of the self to revealed preference theory.

unstable and contested, so some might say this explanatory strategy is overly ambitious. One might then be tempted to despair about making time central to economics, and retreat to the discourse of mathematical equilibrium models that set time and realism aside. However, such a retreat, I suggest, only confirms the openness of the economy since its object is to close our explanations, and one can only close something that is open.

What can we then accomplish in this uncertain scientific environment? I have tried to argue that agents' pursuit of individual identity remains one relatively certain phenomenon. My argument for this turns on whether we believe agents are self-organizing and engage in an adjustment behavior in which they act upon themselves. The motivation for this view was the bounded individuality idea. If individuality in what different forms it takes is the product of action, then it is bounded in a reflexive way by that action, just as in behavioral explanations rationality is bounded in a reflexive way by its own mechanisms. Simon was the original proponent of the bounded rationality idea. But his self-organizing systems idea, once applied to individual agents, allows us to extend the bounded rationality idea to the idea of bounded individuality. This paper aimed to develop that account by framing it explicitly in identity terms on the assumption that any systematic account of economic agents needs to be framed in those terms.

References

Archer, Margaret (1995) *Realist Social Theory: The Morphogenetic Approach*, Cambridge: Cambridge University Press.

Aristotle (1963) *Categories* and *De Interpretatione*, J. H. Ackrill, trans., Oxford: Clarendon Press.

Arthur, W. Brian (1995) "Complexity in Economic and Financial Markets," *Complexity* 1 (1).

Arthur, W. B., Holland, J., LeBaron, B., Palmer, R., and P. Tayler (1997) "Asset pricing under endogenous expectations in an artificial stock market," in W. B. Arthur, S. Durlauf & D. Lane, eds., *The Economy as an Evolving Complex System II*, Addison-Wesley, Reading, MA, pp. 15–44.

Boumans, Marcel (2015) *Science Outside the Laboratory*, New York: Oxford.

Broome, John (2004) *Weighing Lives*, Oxford: Oxford University Press.

David, Paul (1985) "Clio and the Economics of QWERTY," *American Economic Review* 75: 332–337.

Davidson, Paul (1996) "Economic Theory and Reality," *Journal of Post Keynesian Economics* 18: 479-508.

Davis, John (2007) "Identity and Commitment: Sen's Fourth Aspect of the Self," in F. Peter and H. B. Schmid, eds., *Rationality and Commitment*, Oxford: Oxford University Press: 313-335.

Davis, John (2011) *Individuals and Identity in Economics*, Cambridge: Cambridge University Press.

Davis, John (2015) "Bounded Rationality and Bounded Individuality," *Research in the History of Economics and Methodology* 33 (2015): 75-93.

Grűne-Yanoff, Till, Caterina Marchionni, and Ivan Moscati (2014) "Introduction: Methodologies of Bounded Rationality," *Journal of Economic Methodology* 21 (4): 325-342.

Hausman, Daniel (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.

Hausman, Daniel (2012) *Preference, Value, Choice, and Welfare*, Cambridge: Cambridge University Press.

Hausman, Daniel (2016) "On the Econ Within," *Journal of Economic Methodology* 23 (1): 26-32.

Hedoin, Cyril (forthcoming 2016) “Sen’s Critique of Revealed Preference Theory and Its ‘Neo-Samuelsonian’ Critique: A Methodological and Theoretical Assessment,” *Journal of Economic Methodology* 23 (4).

Horst, Ulrich, Alan Kirman, and Miriam Teschl (2007) “Changing Identity: The Emergence of Social Groups,” Princeton, NJ: Institute for Advanced Study, School of Social Science, Economics Working Papers: 1-30.

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden (2016a) “Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioral Welfare Economics,” *Journal of Economic Methodology* 23 (1): 1-25.

Infante, Gerardo, Guilhem Lecouteux, and Robert Sugden (2016b) “On the Econ Within: A Reply to Daniel Hausman,” *Journal of Economic Methodology* 23 (1): 33-37.

Kirman, Alan (2011) *Complex Economics: Individual and Collective Rationality*, London: Routledge.

Lawson, Tony (1997) *Economics and Reality*, London: Routledge.

Lawson, Tony (2013) “Soros’s Theory of Reflexivity: a critical comment,” *Revue de Philosophie Économique* 14 (1): 29-48.

Merton, Robert K. (1948) "The Self Fulfilling Prophecy," *Antioch Review* (2): 193-210.

Rice, Hugh (2015) "Fatalism," *The Stanford Encyclopedia of Philosophy* (Summer Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/sum2015/entries/fatalism/>; accessed March 9, 2016.

Rizvi, S. Abu Turab (2006) “The Sonnenschein-Mantel-Debreu Results after Thirty Years,” *History of Political Economy* 38 (annual supplement): 228-245.

Ross, Don (2005) *Economic Theory and Cognitive Science*, Cambridge, MA: MIT Press.

Ross, Don (2006) “The Economic and Evolutionary Basis of Selves,” *Cognitive Systems Research* 7: 246-258.

Ross, Don (2007) "*H. sapiens* as ecologically special: What does language contribute?" *Language Sciences*, 29.5: 7 10-731.

Rosser, Barkley (2015) "Reconsidering Ergodicity and Fundamental Uncertainty," *Journal of Post Keynesian Economics* 38: 331-354.

Searle, John (2001) *Rationality in Action*, Cambridge, MA: MIT Press.

Sen, Amartya (2002) *Rationality and Freedom*, Cambridge, MA: Harvard University Press.

Simon, Herbert (1955) "A behavioral model of rational choice," *Quarterly Journal of Economics* 69: 99-118.

Simon, Herbert (1956) "Rational Choice and the Structure of the Environment," *Psychological Review* 63: 129-38.

Simon, Herbert (1962) "The Architecture of Complexity," *Proceedings of the American Philosophical Society* 106 (6): 467-482.

Simon, Herbert (1990) "Invariants of Human Behavior," *Annual Review of Psychology* 41: 1-19.

Velupillai, K. Vela (2011) "Towards an algorithmic revolution in economic theory," *Journal of Economic Surveys* 25 (3): 401-430; republished in *Nonlinearity, Complexity and Randomness in Economics*, Stefano Zambelli and Donald George, eds., Hoboken: NJ: Wiley-Blackwell, 2012; 7-35.