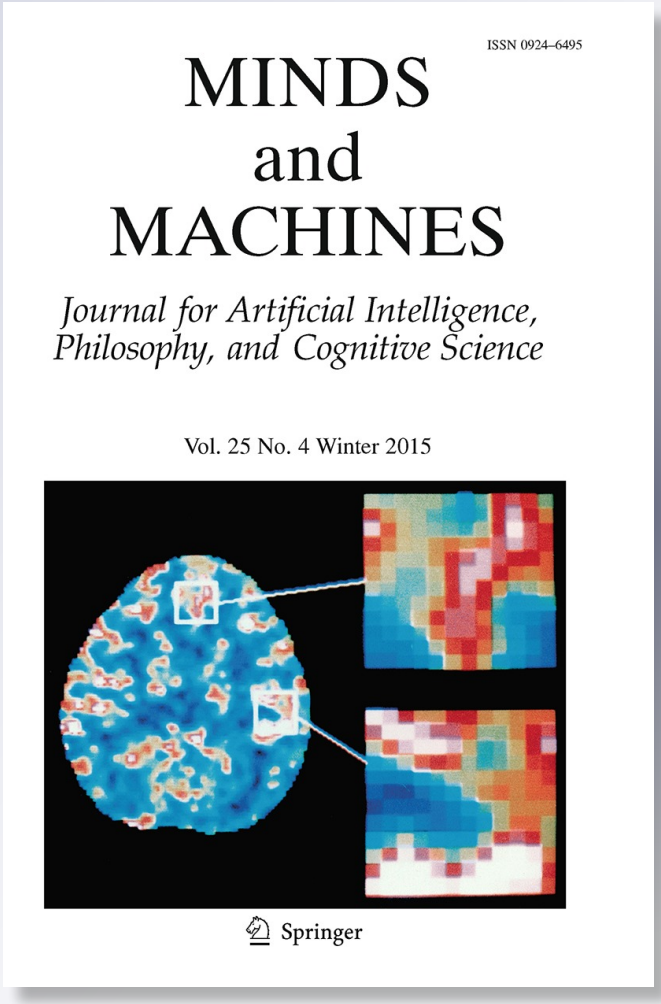
Is Economic Rationality in the Head?

Kevin Vallier

Minds and Machines
Journal for Artificial Intelligence,
Philosophy and Cognitive Science

ISSN 0924-6495
Volume 25
Number 4

Minds & Machines (2015) 25:339-360
DOI 10.1007/s11023-015-9386-6



Springer

Minds & Machines (2015) 25:339–360
DOI 10.1007/s11023-015-9386-6

Is Economic Rationality in the Head?

Kevin Vallier[1]

Received: 3 June 2015 / Accepted: 17 November 2015 / Published online: 28 November 2015

Abstract Many economic theorists hold that social institutions can lead otherwise irrational agents to approximate the predictions of traditional rational choice theory. But there is little consensus on how institutions do so. I defend an economic internalist account of the institution-actor relationship by explaining economic rationality as a feature of individuals whose decision-making is aided by institutional structures. This approach, known as the subjective transaction costs theory, represents apparently irrational behavior as a rational response to high subjective transaction costs of thinking and deciding. The theory has two attractive features. First, it reconciles rational choice theory with the increasing body of evidence cataloguing putative errors in economic decision-making. Second, it vindicates the explanatory power of individual choice against externalist challenges; the subjective transaction costs theory keeps economic rationality in the head.

Keywords Economic rationality · Economic agency · Market rationality · Subjective transaction costs · Behavioral economics · Heuristics and biases · Extended mind

Introduction

Defenders of neoclassical economic theory claim that it has explained a wide range of economic phenomena, but its critics counter that individual human decision-making is subject to systematic rational bias. To resolve the apparent tension, a number of theorists have argued that the structure of social institutions leads agents to approximate the traditional rational actor model employed by neoclassical economics.[1] In other words,

[1] For a helpful review of some of the puzzles raised, see Binmore (2007), especially the introduction.

✉ Kevin Vallier
kevinvallier@gmail.com

[1] Bowling Green State University, Bowling Green, OH, USA

the empirical successes and failures of rational choice theory in economics are best explained by institutional factors. But an institutionalist approach raises a puzzle for economic agency. If institutions are required to produce truly rational choice, how do they do so? In this paper, I defend an *economic internalist* answer to this question. An internalist explains economic agency and rationality as properties of individuals whose decision-making is aided by institutional structures. In this vein, I will argue that treating individuals as if they are economic agents better explains the relationship between individual irrationality and institutional rationality than attempts to deny that individuals are the right unit of explanation.

I will also explore a particular version of internalism that I call the *subjective transaction costs* theory of economic rationality. The subjective transaction costs theory holds that institutions generate rational economic behavior when they reduce the subjective transaction costs required to make rational[2] decisions. Thus, many apparent individual irrationalities can be explained in terms of agents rationally economizing on subjective transaction costs. Effective institutions reduce these costs.

The most serious challenge to economic internalism holds that explaining the variation of economically rational behavior across different institutional contexts requires building institutional facts into the explanatory concept of an economic agent. These economic *externalist* challenges explain economic agency as a combination of the common sense conception of the individual plus his environment. Debra Satz and John Ferejohn (1994) have defended an externalist view, arguing that since institutions are required to produce rational choice, that they should sometimes be understood as *components* of rational choice. Similarly, some cognitive scientists defend the view that many cognitive functions are embodied and even extended into the environment. Cognition does not begin and end with the brain. In response, following concerns advanced by Hausman (1995) and critics of embodied and extended mind approaches to cognition (Adams and Aizawa 2009; Prinz 2009; Rupert 2004), we can plausibly deny that institutional constraints partly constitute economic decision-making. Instead, social scientists do best to explain economic rationality as merely *embedded* in these institutions by making economic rationality *depend* on the institutional structures without partly *consisting* in those structures.[3] Externalist challenges must be answered to buttress economic internalism's plausibility.

Some terminological notes before I proceed. I shall use the term "economic agency" to refer to a unit of explanation with intentional causal power, namely an agent that makes choices that can bring about the satisfaction of her goals. Agents have preferences and beliefs, so when I discuss the extension of agency into the environment, I mean that preferences and beliefs are extended as well. I understand "economic cognition" as the process of generating economic choices, some rational, some not. Further, I take "economic rationality" to be an expression of economic agency that occurs when reasoning conforms to certain canons of good

[2] That is, *faux-rational* decisions, decisions whose cost-benefit analysis excludes subjective transactions costs. I explain the idea of faux-rational choice below.

[3] The embedded-extended mind debate maps onto internalist-externalist interpretations of rational choice theory, as both debates concern whether scientific theories best explain psychological and economic data based on a *choice unit* that is located entirely within the head or partly outside of it.

reasoning, which I explain in "Rational Choice Theory and Its Critics" section. I shall call agents *efficient* when they make preference-satisfying choices, and when they exhibit a high degree of rationality. When I describe economic agency and economic rationality as *embedded*, I mean that they are heavily dependent on institutional factors in order to make effective, rational choices. And when I describe economic agency as *embodied* or *extended* into the environment, I mean that when we individuate economic agents and their choices we include in the unit of explanation the institutions and rules that affect choice in that particular case. I take an institution to be any social order that prescribes individuals or groups to follow particular rules based on a social purpose. For my purposes, I will focus on the fact that institutions are comprised by and impose rules of behavior of those acting within it.

I want to stress here that my thesis attempts to provide an account of successful scientific *explanation*. While I am not concerned primarily with social ontology, and my focus on explanation may have ontological implications, explanation is my primary aim. I am asking whether it makes more sense to characterize the individual agent as a unit of explanation as a commonsense "in the head" agent, or as an agent or agents constituted by a superset or subset of the brain. That is, how should behavioral scientists carve up the world in explaining the apparent discrepancies between the relevant literatures in psychology and economics?

I believe that giving an account of the extent, or perhaps *extension*, of the economic agent is necessary to clarify and potentially resolve three contemporary scientific and philosophical controversies. First, an account can help to resolve controversies within cognitive science over whether cognitive psychological explanation should appeal to an embodied or extended mind. If economic rationality, a very important part of agency, can be given a compelling internalist explanation, that helps undermine one form of argument for the extended mind. Second, an account of the extension of the economic agent can help to resolve the ongoing controversy in the philosophy of social science, and philosophy of economics in particular, about the right level of social scientific explanation, that is, whether economic agency should be understood in an internalist or externalist fashion. Finally, I believe an attractive internalist account of economic agency can help to resolve the increasingly sophisticated debate among experimental economists about how individual economic agents come to approximate the rational actor model in experimental settings. An internalist account will help to explain why individuals are able to adapt to some institutional rules and not others. These problems are further resolved by my defense of the subjective transaction costs theory, so that a fleshed out version of internalism can function as a contrast to externalist positions; while in my opinion Hausman (1995) has pointed the way towards answering moderate externalists like Satz and Ferejohn, we still need further work on a positive internalist alternative.

This essay proceeds in five parts. In "Rational Choice Theory and Its Critics", I briefly review rational choice theory as employed by economists and empirical challenges to it. In "Attempts at Reconciliation", I criticize some relatively simple attempts to account for the successes and failures of rational choice theory in different institutional contexts. "Externalist Challenges" examines two externalist

challenges to internalism that attempt to explain economic agency in terms of individuals plus institutional constraints. While the empirical data strongly suggest that economic rationality often proceeds by means of embedded cognition, I argue that they do not require the stronger claim that economic cognition is embodied in the environment, as externalists maintain. "The Subjective Transaction Cost Theory: An Internalist Alternative" develops and defends the subjective transactions costs theory of economic rationality. The concluding section reflects on how to defend a particular version of economic internalism.

Rational Choice Theory and Its Critics

Most economists employ rational choice theory to generate economic models.[4] Individual action is said to be rational when an agent's choices are consistent with an ordinal utility function with weakly ordered preferences, which must be complete, reflexive and transitive.[5] According to expected utility theory, in cases of decision-making under risk, rational actors combine the utility of an outcome with the outcome's probability and select the greater sum of the products of utility and probability over the lesser.[6] Economists add conditions to the rational actor model to generate the familiar *homo economicus* model of economic behavior.[7] First, homo economics prefers more to less. Second, homo economicus makes choices at the margin. Homo economicus is also often assumed to have wealth-maximization preferences, though this is not essential to it. In this paper, we can examine the rational actor model in economics without taking on assumptions about wealth maximization, but I do assume that rational actors prefer more to less and make choices at the margin.

The rational actor model has been widely criticized.[8] Cognitive psychologists, such as Kahneman (2013) often in conjunction with Tversky (Kahneman and Tversky 1974), and economists such as Thaler (1994) have documented voluminous cases of poor human decision-making such as the conjunction fallacy (Kahneman and Tversky 1982, 92). Studies have consistently shown that individuals make poor decisions when conjoining identical probabilities and utilities with divergent descriptions. Related phenomena are framing effects, where the description of a choice scenario significantly influence the choice made. Framing effects seem to

[4] This said, decision theory is employed differently by different subfields in economics. For more on the plural understanding of rationality used by economists, see Cowen (2004). Our scope is restricted to those areas of economics that employ the theory in relatively unmodified forms; to the extent a sub-field of economics employs traditional decision theory, to that extent what I say here applies to it.

[5] I draw on the list provided by Gaus (2007, 36–7). But for classic treatments, see Luce and Raiffa (1989), Savage (1972). Indifference is typically symmetric and transitivity can be weakened to quasi-transitivity.

[6] For a clear, concise explanation of expected utility theory, see Resnik (2000, 45–118).

[7] The following list of features of homo economicus is given by ibid., 19–27.

[8] Many anthropologists have argued that in traditional societies, economic decisions are governed by norms of reciprocity. See Polanyi (2001), Sahlins (1972), esp. 1–40. Herbert Simon's work on bounded rationality maintains that human rationality is subject to certain cognitive limitations. See Simon (1984).

imply a reversal of preference, and preference reversals challenge the rational actor model's assumption of complete and intransitive preferences; these experiments appear to show otherwise because they seem to demonstrate that preferences are often inconsistent because they're easily reversed. Empirical evidence of these effects has generated the "heuristics and biases" tradition in cognitive science. According to this view, humans have evolved to process information according to simple cognitive heuristics with low cognitive costs and that produce systematically biased behavior. Despite promoting survival, these heuristics mostly get the answer right (Kahneman and Tversky 1973) but often err (Bower 1996; Gould 1992).

All this is familiar, but bears repeating.

Attempts at Reconciliation

Behavioral Economics and Doubling Down on Irrationality

One common view among economists is that the irrationality of individual actors is irrelevant so long as models have predictive power. This methodological disposition derives from Friedman's (1953) article "The Methodology of Positive Economics," where he defends the vaguely positivist thesis that a hypothesis matters "if it abstracts the common and crucial elements" from a range of phenomena and generates predictions.[9] I will assume, however, that, most social scientists prefer a genuine reconciliation between the heuristics-and-biases tradition and microeconomic analysis.

One increasingly popular response to claims that there is a tension between empirical work demonstrating individual irrationalities and other work documenting rational or efficient action in institutional contexts is simply to deny or downplay the latter. Some behavioral economists and cognitive scientists pursue this line, arguing that economic evidence for rationality in market contexts is ambiguous and illusory. For example, in *Predictably Irrational,* Dan Ariely argues that humans are often fooled into thinking that they make rational decisions when they do not. Ariely (2008, 26–8) has documented statistically significant anchoring effects even for MIT-trained MBA students. Ariely concludes that contrary to the economist's assumption that prices are determined by supply and demand, "what consumers are willing to pay can easily be manipulated" (46).

Ariely omits any discussion of decades-old experimental work in economics which attempts to verify economic models by running economic experiments with interacting test subjects. Traditional behavioral economists like Ariely focus more often on individual and group behavior isolated from controlled iteration. Vernon Smith, perhaps the pioneer of this form of experimental work, has conducted hundreds of experiments that replicate market efficiency, where actors behave rationally by maximizing utility in experimental contexts.[10] In these experiments,

[9] Obviously there are many positivisms. For our purposes, I am simply examining the surprisingly still popular Friedman, Popper-based positivism commonly accepted by economists.

[10] For a summary analysis of Smith's work and the broader conception of economic rationality he endorses, see Smith (2008). Smith has been the main articulator of the idea of "ecological rationality" that shares much with the STC theory I defend.

actors engage with one another in a serial manner, repeating their interactions over discrete periods of time. Smith has found that subjects become gradually more effective at making utility-maximizing choices. That is, they are increasingly able to make all utility-increasing trades. While Smith notes that his experiments do not always confirm standard expected utility theory, "experimental tests of market theories, which explicitly assume expected utility (or value) maximization, have *not* falsified many of these theories" (1985, 267). And while many initial tests show marked irrationalities, over time behavior approximates economic models. Smith has long argued that, "markets may induce greater 'rationality' in behavior because they force or promote a response to, or discovery of, opportunity cost conditions, that need not be readily forthcoming when agents merely think about the choices they make." Framing effects "do relatively low level damage to [the expected utility hypothesis]" for economic actors can learn to avoid these effects (268-69). Economic actors do not become expert statistical reasoners. Rather, they learn to avoid options that lead to net cost. Experimental economists like Smith freely admit that individuals exhibit framing effects and that, initially, groups do as well. The question is whether these effects will persist over time as individuals become more familiar with how to avoid net costs. Smith and his colleagues reject Ariely's conclusions because much work in behavioral economics ignores the data demonstrating that iterative exchanges reduce anchoring effects, among others. Behavior consistent with rational choice theory will often occur when agents interact under the appropriate rules, even if they act irrationally when unconstrained by those rules.

Like the positivist solution, the approach of behavioral economists like Ariely discounts the data rather than grappling with the tension. Individuals are subject to deep irrationalities and yet engage in rational behavior in iterative social interactions. The tension cannot be denied; it must be resolved.

Alternative Expected Utility Theories

Some of the more sophisticated attempts to reconcile irrational psychologies with rationality in interactive contexts documented by experimental economists deploy alternative expected utility theories. These theories depart, in one way or another, from the standard assumptions of expected utility theory outlined above. The two types of alternative expected utility theories (henceforth, EUT) are "conventional" EUTs and "nonconventional" EUTs.[11] Conventional EUTs try to preserve expected utility theory while relaxing some of its standard assumptions. One conventional EUT is weighted expected utility theory, where outcomes of actions are "weighted" when they are given more value in an expected utility calculation vis-à-vis other outcomes that exceeds the value warranted by the product of the utility and probability of the outcome.[12] Nonconventional EUTs or "procedural" theories only

[11] These terms derive from Starmer (2000).

[12] One illustration of weighted utility theory can be found in Chew and MacCrimmon (1979).

assume that economic actors use decision heuristics.[13] The most prominent nonconventional expected utility theory is cumulative prospect theory (Kahneman and Tversky 1992). In cumulative prospect theory, cumulative probabilities, rather than individual probabilities, are built into choice options.[14] Consequently, real agents make choices in accord with the probability of a class of actions of which the relevant choice is member. Prospect theorists then argue that the irrationally high weights we assign to low probability events result from the decision-making phase where the class is delineated. Heuristics are used to determine which events are risky, and though they are cognitively cheap and evolutionarily successful, such heuristics generate irrational behavior in many cases.

Unfortunately, both conventional and nonconventional expected utility theories were formulated somewhat independently of (and arguably prior to) serious engagement with the literature on repeat interaction. In the review of Vernon Smith's work, we saw that feedback from repeated social interactions helps actors in single-shot bets approximate the traditional rational actor model, though whether, say, the loss aversion postulated in prospect theory, tends to diminish is highly context sensitive (Smith 2008, 149–156). Alternative EUTs merely describe utility functions that reflect the widespread effects of heuristics and biases on economic behavior; experimental evidence suggests that whether they accurately describe agent behavior in market contexts depends on a host of environmental factors. Since our aim is to determine how institutional rules interact with cognitive structures over time, alternative EUTs are only part of the story.

Externalist Challenges

What Externalist Challenges Share

Before developing a particular internalist account of economic agency, we must justify the conclusion that economic rationality is "in the head." To do so, I must address what I shall call *externalist* challenges to the view that economic rationality and agency are best understood as properties of individuals. All externalist challenges claim that the best way to explain how institutional rules aid economic rationality is to extend economic agency into the body and the environment. In some sense or another, economic agency is not confined to the head. The proper unit of social scientific scrutiny, therefore, includes more social and physical structures than the brain and its sensory inputs and outputs. Economic agency and rationality

[13] This distinction is not well-formed. Heuristic use can be represented as optimizing a preference function. The preference function is simply a model for representing choice; utility theory does not require psychological underpinnings. See Gaus (2007, 1–6).

[14] Some claim that Kahneman's work suggests the more radical thesis that human decision-making cannot be characterized in terms of preference orderings and that such orderings are formed on the fly. But, importantly, Kahneman and his co-authors have not generally taken this more radical step. Instead, they have tried to develop an economic model that understands these phenomena as consistent with certain sorts of orderings, orderings that are often quite stable given persistent heuristics and biases.

are external, then, when social scientists individuate them to include institutional rules and constraints as part of the unit of explanation we typically refer to as an economic agent or chooser.

To define the externalist challenges more precisely, we can appeal to a distinction between two theses in cognitive science, what Robert Rupert calls the "hypothesis of extended cognition" and the "hypothesis of embedded cognition." (Rupert 2004, 389–393).[15] The hypothesis of extended cognition holds that "human cognitive processing literally extends into the environment surrounding the organism, and human cognitive states literally comprise—as wholes do their proper parts—elements in the environment." In contrast, and more modestly, the hypothesis of embedded cognition holds that "cognitive processes depend *very* heavily…on organismically external props and devices and on the structure of the external environment in which cognition takes place." Externalist challenges hold that economic rationality is extended, whereas I shall argue that economic rationality is merely embedded. Thus, externalist challenges deny that we can resolve the tension between individual irrationality and market-clearing behavior by relying on individual agents as the proper unit of explanation. In one form or another, all externalist challenges push social scientists to explain much of market-clearing behavior in terms that extend beyond the common sense individual economic agent.

We should address the concern here that the debate between internalists and externalists is unimportant because it has no explanatory consequences for economic science. This is because an extended agent will arguably have a different utility function than a merely embodied agent. If we model an agent as extended into the world, then in many cases simple descriptions of his choices will have to change in ways that are arguably not extensionally equivalent to internalist models. For instance, an extended agent will have preferences over states of affairs that will not include choosing under different institutional conditions because it is *constituted by* those conditions, and would not exist but for them. In contrast, the embedded agent will have preferences over states of affairs that include varied institutional conditions. Whether this difference turns out to be empirically consequential is an open question, but it is plausible that the difference could be significant.

So the internalist-externalist debate does have consequences for economic science. Consequently, externalist challenges require refutation if I am to develop a plausible internalist resolution of the tension between individual irrationality and market-clearing behavior and between the conflicting data of economics and cognitive science. For this reason, I will examine two externalist views that, while not accounts of specifically economic agency, can be employed to challenge my form of economic internalism: the extended cognition thesis defended by Andy Clark and the "moderate" social scientific externalism defended by Satz and Ferejohn. But I must first review data that both challenges can draw on.

[15] Critics of extended mind views follow Rupert in drawing a distinction between institutional causation and institutional constitution. For discussion, see Wilson and Foglia (2011), esp 6.3.

Theoretical and Empirical Support for Externalist Challenges

Externalists can use work from experimental economics and new institutionalist economics to justify their inference from the limited rationality of individual agents to the extension of agency and choice into the environment. They can claim, in short, that the best explanation of economically rational behavior in institutional contexts is that cognition sometimes includes the body and the environment.

New institutional economists Arthur Denzau and Douglass North (1994) argue that neoclassical economics relies on a conception of "substantive rationality" or rationality that is contained entirely in the head (6). But economic and political history cannot be understood in terms of individuals acting in accord with standard rational choice theory. Instead, institutions are required to reduce transaction costs, thereby making market efficiency possible. Ronald Coase was one of the first economists to focus on the role transaction costs in economic behavior. Transaction costs are "the costs other than the money price that are incurred in trading goods or services," that is, the costs of participating in a market.[16] Standard neoclassical economics is said to assume zero transaction costs, assuming "frictionless" exchange. Following Coase, Denzau and North argue that some economic phenomena cannot be explained without appealing to the impact transaction costs have on market functioning. For instance, Denzau and North apply this Coasean framework to economic history, arguing that ideologies and cultural institutions arise to reduce transaction costs as well. Economic and social institutions are also mechanisms that "offload" cognition onto the environment, or in their terms, reduce transaction costs.[17] Once transaction costs are sufficiently reduced and economic cognition is sufficiently embedded, choice options are limited and rational decision-making follows. North developed his view partly in light of his observation that markets performed poorly when implemented in the former Soviet Union. He argues that markets often failed there because the Soviet republics lacked the cultural and moral institutions required to sufficiently reduce transaction costs. Individuals could not trust one another and did not know how to run businesses (North 2005, 146–170).

Denzau, North, and Clark (1996) also cite experimental evidence that individual rationality is less important for market efficiency than is often thought. Their primary evidence comes from economic experiments where human traders are replaced with "zero-intelligence" programs. These programs submit random bids and offers subject only to a budget constraint—they engage in no explicit economic reasoning. Dhananjay K. Gode and Shyam Sunder find that "imposing a budget constraint (i.e., not permitting traders to sell below their costs or buy above their values) is sufficient to raise the…efficiency of these auctions close to 100 %" (Gode and Sunder 1993, 119).[18] They go on to claim that the efficiency of markets appears to be due to "its structure independent of traders' motivation, intelligence, or learning" (Ibid.). Again,

[16] For a brief explanation of transaction costs, see: http://www.auburn.edu/johnspm/gloss/transaction_costs.

[17] One need not be an economic determinist to appreciate this point.

[18] Also see Gode and Sunder (1997).

the shape of institutions produces rational action.[19] Like North and Denzau, they find that institutions are required to make markets work. Smith (despite endorsing internalism) has argued that limited information and reliable trading rules "are sufficient to produce competitive market outcomes at or near 100 % efficiency" (Smith 1982a, 167). But Gode and Sunder go farther: economic actors do not need much information *or rationality* in order to produce efficient outcomes under the right rules. As they say, "Adam Smith's invisible hand may be more powerful than some have thought; it can generate aggregate rationality not only from individual rationality *but also from individual irrationality*" (Gode and Sunder 1993, 119).

Given how little "in the head" rationality matters for efficient economic institutions and efficient actor behavior, we might think that the individual agent is no longer the best unit of explanation. Perhaps we do better to model economic cognition as at least sometimes extending into those institutions.[20]

Clark's Externalism and Embodied Cognition

The externalist challenge is that economic cognition is best explained as extending into the institutions in which economic agents are embedded. In other words, the institutional rules that enable more effective economic decision-making partly constitute economic agents. Clark is well known for defending the view that much cognition is "embodied" in the environmental and institutional arrangements with which agents typically interact (Clark 1997, 2008). For Clark, cognition arises from the interplay between brain, body and environment. Human cognition is unique not because humans are good individual reasoners but because they possess "amazing capacities to create and maintain a variety of special external structures" (179). He continues: "these external structures function so as to complement our individual cognition profiles and to diffuse human reason across wider and wider social and physical networks whose collective computations exhibit their own special dynamics and properties" (Ibid.). Cognition extends into the environment, making complex forms of cognition easier, by "dissipating" some forms of reasoning, increasing the brain's ability to execute increasingly complex plans.

The case for extending Clark's externalism about cognition to economic rationality is rooted in the observations, listed above, that economically efficient choices are made by (traditionally individuated) actors that have extremely limited intelligence and that frequently make irrational decisions when left to themselves.[21] Economically efficient outcomes are correlated with institutional structures more

[19] Their claim resembles Alchian (1950), which sees profits as a constraint rather than an objective. Since only firms that make profits survive in the marketplace, firms may *appear* to maximize profits when they have other aims, or none at all. I thank Mike Munger for this point.

[20] There is an additional argument for externalism that, given how much institutional factors bear on *what* agents prefer, we should extend the mind into the environment in order to generate preference stability. But I take it such an argument can only be motivated if we already have an account of what agents want in hand, which would require having already settled the matter of dispute in this paper. I thank an anonymous referee for raising this point.

[21] Clark has, however, applied much of his theoretical framework to understanding the successes of neoclassical economic theory despite its putatively implausible conception of decision-making. See Clark (1996).

than they are to individual beliefs and preferences. In other words, efficient decision-making is coupled with institutional rules rather than individual mental states. This is sometimes described as a "coupling-constitution" argument, which involves an inference from the fact that a feature of the environment is somehow causally tied to a cognitive process to the claim that the environmental process is part of the cognitive process.[22] A "coupling" argument demonstrates that two phenomena are tightly empirically correlated, whereas a "constitution" claim construes them as part of the same unit of explanation. We can build an analogous argument against economic internalism by arguing that the empirical "coupling" of economic rationality and institutional rules should lead us to explain economic agents as constituted by those institutional rules.

The trouble with coupling arguments, as critics have repeatedly pointed out, is that they appear to rest on what Fred Adams and Kenneth Aizawa have called the "coupling-constitution fallacy." One commits the fallacy by relying solely on the claim that an environmental process is coupled with a cognitive process in order to demonstrate that the environmental process partly constitutes the cognitive process. The argument is fallacious because coupling is insufficient to demonstrate constitution. Critics of embodied cognition approaches have argued that most if not all of the arguments for embodiment depend on this fallacy. Something similar holds in our case, as nothing in the data reviewed requires going beyond a coupling relation between the environmental and individual cognition to a constitution relation. If so, we can set the radical externalist challenge aside.

However, several philosophers and social scientists have recently rejected Adams and Aizawa's use of the coupling-constitution fallacy. Menary (2012) points out that Adams and Aizawa may have misunderstood the form of arguments for the extended mind. The aim of the extended mind theorist "is not to show that artifacts get to be part of cognition just because they are causally coupled to a preexisting cognitive agent, but to explain why X and Y are so coordinated that they together function as Z, which causes further behavior" (12). Clark makes a similar point in arguing that the goal of extended mind arguments is not to make some external object "cognitive" but instead to "make some object, which in and of itself is not usefully…thought of as *either cognitive or* noncognitive, into a *proper part of some cognitive system,* such as a human agent" (Clark 2012, 83). The externalist is not arguing that, because, say, economic institutions are coupled with rational economic choice, that they should therefore be included in the description of a preexisting rational agent, but rather that because economic institutions and rational individual economic choice are coupled, that they are therefore part of a real entity Z that engages in rational choice.

In my view, the best way to adequately respond to Clark is to deny that the evidence presented above requires postulating some supra-skull Z-type agent, whose description includes the brain and economic institutions. Clark need not be guilty of a coupling-constitution *fallacy*, but if there is an internalist alternative to the extended mind, our ordinary commonsense notion of "in the head" agency can

[22] See Adams and Aizawa (2009) for discussion.

be safely preserved. That is, we need not adopt more counteruitive externalist accounts unless we're forced to do so.

A potential response to my approach could follow Don Ross and James Ladyman's recent criticism of Adams and Aizawa (Ross and Ladyman 2012). On Ross and Ladyman's view, the coupling-constitution fallacy isn't a fallacy in scientific reasoning because contemporary science doesn't appeal to this "metaphysical" distinction. In fact, they claim "metaphysical considerations should play no role in deciding how to model cognition" (156). Instead, "modelers will and should draw system boundaries in whichever ways maximize efficient capture of local phenomena." Further, in "mature" sciences, "the intuitive distinction between causes and constitution…tends to be abandoned as sciences converge on robust models of general structure" (157). As a result, the coupling-constitution distinction "is an abstraction that does not correspond to any general idea that figures non-metaphorically in science," which should cast "strong doubt" on whether the coupling-constitution distinction should be allowed into scientific explanation. Ross and Ladyman claim not only that the distinction is not present in physics, but that *economists* make no use of it either. The distinction between an agent and its environment is "model relative" (163). That is, systems are distinguished "by reference to variables codetermined as a set *given* simultaneous identification of some other set of variables as exogenous." Economists widely believe that "interesting phenomena admit of multiple parsings along different endogenous-exogenous boundaries." So the challenge to Adams and Aizawa is not that the extended mind thesis is *true*, but rather that their argument against the extended mind fails because the coupling-constitution fallacy does not figure into mature sciences and so should not figure into the operation of cognitive science. For our purposes, Ross and Ladyman can be interpreted as claiming both that the coupling-constitution fallacy is no mark against extending economic rationality outside of the head and that there is no presumption within economic science on behalf of our commonsense intuition that agency is "in the head."

Ross and Ladyman figure into the externalism-internalism dialectic by undermining a central argument against externalism, namely arguments based on the identification of the coupling-constitution fallacy. However, in doing so they also deprive the extended mind theorist of her own positive arguments. Externalists encourage the misuse of the coupling-constitution "metaphor" in a fashion that is ultimately "counterscientific" (164). So even if Ross and Ladyman are correct that the coupling-constitution fallacy is not a fallacy at all, they deprive us of reason to pursue externalism in the first place.[23]

[23] One reason Ross wishes to undermine the use of "metaphysics" of the sort putatively appealed to by the externalist is that he wishes to advance his own account (Ross 2005) of economic rationality where a commonsense agent is understood as a coalition of sub-agents, each of whom tend to follow microeconomic laws. Ross claims that "People are politically complex societies of temporally located selves" (186). The problem with Ross's view is that it expressly requires dismissing "the explanatory value of supposing that the concept of qualia picks out any theoretical unified class of cognitive objects" (231). Denying the usefulness of the idea of qualia, however, ignores data that we can use to choose between social scientific theories, namely conscious experience. Denying the importance of subjective mental properties in social scientific explanation is deeply implausible, so I set his view aside here.

Satz and Ferejohn's Externalism

We might opt for a more modest externalist challenge in reply. Satz and Ferejohn (1994) defend "moderate externalism" about social scientific explanations generally. On their view, the internalist holds that social scientific events can only be explained by intentional states interior to the individual agent. Satz and Ferejohn claim that while individual mental states are relevant to social scientific explanations, their connection is "remote." Consequently, moderate externalism "does not explain behavior in terms of these mental entities; it merely shows that behavior can be interpreted as consistent with them" (77). It explains social scientific events in terms of the "structure in which [individual action] is embedded" (78). Individuals are not always the basic unit of analysis for the moderate externalist. Instead, sometimes the right unit of analysis is a combination of individuals and institutional conditions. It is important to note that, in contrast to some other externalists, Satz and Ferejohn recognize that "for many purposes, an internalist, agent-centered perspective on action is crucial." Thus, externalists differ concerning *how often* they propose setting aside internalist explanations. Clark and others think that the nature of cognition is such that it can routinely be characterized as extended. But Satz and Ferejohn only wish to depart from internalism in restricted cases.

Satz and Ferejohn argue that while one can always give an internalist explanation of a particular choice phenomenon, externalist explanations have benefits. For instance, internalist explanations supposedly have difficulty explaining how social systems can reach equilibrium despite a diversity of "microlevel psychologies." Instead, it is the "structural conditions" that constrain psychological possibilities and "select for compatible microlevel psychologies" (80–81). Satz and Ferejohn's main illustration is the attempt to explain why energy prices are higher in Europe than in the United States. The internalist will explain the difference by pointing to differences in the individual psychologies of managers and customers. The externalist argues that appeals to individual beliefs and preferences can be dispensed with as causal explanations for market outcomes, as there are many psychological make-ups that could lead to these price differentials. In contrast, the range of institutional structures that could produce these outcomes is considerably lower. Satz and Ferejohn could bolster their case for externalism by appealing to the data summarized above. A market can exhibit a high degree of efficiency (all Pareto improvements have been made) even given individual *irrationality*, not to mention a wide array of rational preferences and beliefs. If so, then perhaps the individual is the wrong unit of explanation. We might do better to analyze the individual plus her environment as making rational choices the result in efficient outcomes.

Hausman (1995) has criticized externalism on the grounds that internalism about social scientific explanations is consistent with the data that Satz and Ferejohn discuss. Hausman claims that "nothing in internalism conflicts with [their] observations about the importance of structural relations." (101). Instead, all that Satz and Ferejohn have shown is that the beliefs and preferences of agents are in some cases uninteresting and trivial. There is no reason internalist structural

explanations "cannot complement" psychological explanation by explaining how structural factors make psychological facts causally efficacious (102).[24] The only thing internalism cannot permit is for "social theorists to disregard the empirical difficulties to which their claims about individual psychology may be subject." Internalists only need to claim that externalism is false in order to preserve their claim that the "in the head" individual is the right unit of social scientific explanation.

The rejoinder to Hausman involves insisting that the data motivate a move away from internalism. If efficient market outcomes are robust with respect to individual psychologies, it is easier to justify assigning institutions a constitutive role in social scientific explanation. Thus, the internalist needs to give a plausible account of individual psychological states in cases where their states seem to make little difference with regard to economic outcomes. In other words, we need a well-developed internalist alternative to moderate externalism in order to justify the claim that we can best explain economic behavior by appealing to the in the head individual as the correct choice unit.

The Subjective Transaction Cost Theory: An Internalist Alternative

The Subjective Transaction Cost Theory

In this section, I defend the subjective cost of transacting (STC) theory as an alternative to economic externalism (Marschak 1968; Smith 1982b, 934). Specifically, I advance the STC theory as an explanation of the compatibility of irrational or non-rational actors and efficient market outcomes. Subjective costs of transacting are the costs of "thinking, calculating, deciding and acting" (Smith 1985, 268). These costs are subjective because they are *cognitive,* or internal to an agent's cognitive systems. Further, these are costs that are, to a limited extent, under the agent's control. That is, they can be *paid* in terms of time spent and cognitive energy exerted on the task at hand rather than alternatives. To illustrate, imagine that John wants to buy Wheat Thins at Food City. When he arrives, he finds two differently sized boxes. The smaller box is cheaper but contains fewer Wheat Thins. Typically the large box has a lower unit price, but today the small box is on sale. John must now calculate the relative unit prices of the two boxes himself.[25] For John, division with decimals and multiple digits is a chore. Even with a mobile phone calculator, John must still slow down and delay moving on to other tasks. The unit price calculation is a subjective transaction cost. Such costs are (a) cognitive and (b) volitional.

External structures in the economic environment can reduce the subjective costs of transacting. For instance, shared behavioral expectations can provide the

[24] We could also describe Hausman's claim as that structural explanations can help explain coupling without postulating a constitution relation.

[25] Suppose that your rate of Wheat Thin consumption is an exogenous variable (you won't eat them faster if you have more).

groundwork for cooperation even if people rarely recognize that they do so. These behavioral expectations may help reduce the stress of trusting others. Extending our example, if John worries that his box of Wheat Thins might contain a small piece of metal from a faulty manufacturing plant, he will have to risk buying the box and checking for himself. But if John lives in a developed nation with good institutional rules, such as contract laws, regulations and appropriate informal moral norms, he can effectively eliminate the risk. So the costs of rational economic decision-making fall.

The STC theory understands economic agents as individual persons who economize on subjective transaction costs. That is, agents are instrumentally rational with respect to their total cost. When subjective transaction costs (STCs) are low, actors will appear rational as predicted by standard rational choice theory. However, when STCs are high, actors will appear irrational in the sense that, *excluding STCs*, the agent fails to maximize preference satisfaction. She is unable to make "available" utility gains if we understand availability independently of STCs. In virtue of this inability, she may be counted "irrational." Let us call an agent "STC-rational" when she maximizes preference satisfaction with subjective transaction costs built into her utility function. Let us say that she is "faux-rational" when she maximizes preference satisfaction independently of STCs.[26] The thesis I advance is this: individual faux-irrationality is STC-rationality with high STCs. The STC theory predicts that actors will be faux-rational when STCs are low; faux-rational market behavior should increase within institutions that reduce STCs.

Advantages of the STC Theory

The STC theory is consistent with the insights of the embedded cognition and new institutionalist approaches, which both emphasize that cognition is eased under good institutions. For North and Clark, John's cognition can be offloaded onto the environment when he need not calculate the risk that he will find metal in his Wheat Thins or when unit prices are displayed in the grocery store. The STC theory can also make sense of increased market rationality through repetition. Repetition reduces STCs via social learning under stable institutional rules, which in turn reduces the STCs of acquiring, storing and accessing information. Agents learn to make good decisions through repetition, observing outcomes and adjusting their behavior accordingly. Consider the insights of experimental economics: as trials are repeated, individual behaviors approach efficient market outcomes. If repetition lowers STCs, then it is no surprise that repetition increases rational market behavior. It is no coincidence that Vernon Smith defends a subjective transaction costs theory to explain how repetition produces market efficiency.

[26] I use the term "faux-irrationality" because it seems appropriate to use the term "rational" to apply to an agent's behavior given her total costs of achieving an outcome, including the costs of figuring out that the outcome is good or best. The problem with much of the cognitive psychology and behavioral economics literature is that they charge individuals with irrationality independently of a crucial set of costs, STCs.

The STC approach also renders rational choice theory consistent with the extensive use of heuristics observed by cognitive psychologists. Explicit and calculative thinking is costly even with the relevant information collected, and information gathering has costs of its own. In many cases, heuristics economize on cognitive exertion because they are cognitively cheaper to execute, even if they are less reliable in general (and even this is sometimes doubtful). That is, it can be STC-rational to employ a heuristic when it increases one's expected utility, understood as including STCs. On this view, purported irrationalities documented by behavioral economists are rational responses to high subjective transaction costs.[27] Heuristic use is a rational response to expensive cognition.

Let us illustrate with a well-documented example. There is neurological evidence to suggest that calculation in more explicitly cognitive, less emotional parts of the brain is slower and more cognitively expensive than other forms of neurological processing (Samuels and Stich 2004). Richard Samuels and Stephen Stich therefore defend a *dual processing* theory of reasoning, which has gained considerable prominence over the last decade (Kahneman 2013; also Evans 2003; Sloman 1996). The dual processing theory holds that agents make decisions via two distinct neurological systems. One system is "fast, holistic, automatic, largely unconscious, and requires relatively little cognitive capacity" (Samuels and Stich 2004, 24). The other system is "relatively slow, rule based, more readily controlled and requires significantly more cognitive capacity" (Ibid.). Psychologists who adhere to this theory argue that the former system is evolutionarily older than the latter and that the latter system can be deeply affected by cultural and institutional factors. Differences in aptitude will affect the degree to which the latter, newer system functions adequately, while the older system is hard to improve and is equally error-laden across different agents. While it may be an exaggeration to say that agents have a full-blown *choice* as to which system to rely on, they may remain implicitly responsive to whether using the fast or slow system has greater expected utility. They can use their "fast, holistic, automatic" system that requires "little cognitive capacity" or spend the resources to employ their "relatively slow, rule-based, more readily controlled" system of reasoning. When the information required to make an explicit calculation is costly to collect or when the required calculation is cognitively expensive, it is STC-rational to select less accurate systems of reasoning given the relative costs and benefits of getting the right answer. On occasion agents can employ their slower, system of reasoning to override the outputs of the faster, system. In many cases, agents cannot prevent the older system from operating, but they can prevent its outputs from determining action.

In the same way, when a decision is cognitively inexpensive, i.e., when the subjective transaction costs of calculating are low, rational agents should reduce their heuristic use, and replace heuristics with more accurate processes. In this way, decision-making can approach the axioms of rational choice theory by appealing to explicit, formal reasoning. As seen above, experimental economists have amply

[27] My claim resembles Gigerenzer and Goldstein (1996), though they think classical decision theory is no longer the appropriate *normative* standard of rationality, as their work shows that more "correct" answers can be achieved with algorithms that depart from traditional models. I am not offering a normative standard, just an explanatory theory.

documented this phenomenon. Many experiments that uncover irrationalities implicitly assume that rational action requires complex probability calculations, such as in studies documenting framing effects. But individuals may not realize that they would do better to calculate rather than let their natural heuristics dictate their decisions, such that the lack of realization could be based on an STC-rational economization on costs of cognition.[28]

A critic could argue that it is hard to see how the STC theory can explain errors in preference reversals and errors in probabilistic reasoning, like base-rate fallacies, as a form of maximizing. How, for instance, are we to characterize an agent as rational who reverses his preferences based on framing, or who acts on incorrect probability judgments? In both cases, we need to recognize the difficulty in uncovering the information that would lead one to resist preference reversals or base-rate fallacies. Because persons have natural heuristics that help them frame decisions and judge the likelihood of certain interactions, the STCs of relying on the heuristics are low. Heuristics are how we reason *fast* (Kahneman 2013). To overcome these errors, agents both need more information and the capacity to override their motivation to follow the better information rather than their heuristics. So an agent can be represented as maximizing expected utility by relying on a fallible frame or incorrect probability estimates because she relies on the decision process with lower STCs. Correcting for these phenomena can be costly for agents. To avoid a base-rate fallacy, for instance, an agent must *reason slow* if she is to determine whether she is misusing base-rate information. Even people with complex statistical knowledge, as we've seen above, continue to exhibit bias in those areas. This suggests that the STCs of avoiding these fallacies are high, since they involve information collection, probabilistic calculation, *and* suppressing one's natural motive to render a fast and frugal heuristic-based answer.[29]

[28] Gigenrenzer and Brighton (2009) have argued that heuristic use is not justified by approximating more complex, and so more accurate, decision-mechanisms, but rather as attempts to avoid *variance* errors by relying on bias to make rough and ready calculations that themselves are often more accurate than more complex processes. So it may be mistaken to conceive of the rationality of heuristic use as giving up accuracy in comparison to making complex, explicit calculations. If so, then we can understand heuristic use as a method of becoming more accurate generally, if not more accurate than explicit, complex cognition.

[29] I thank two anonymous referees for raising this objection. One of the referees also raised the concern that the STC view implies that choice in the absence of social institutions will fail to maximize utility, which does not fit well with the view that people are economic agents whether or not they act within social institutions. In response, I would ask what is involved in an agent making choices outside of institutions. The vast majority of human beings are never in this situation. All of their choices will be affected by the social practices and institutional rules that surround them. Even people who live solitary lives have at least been reared within institutional structures whose behaviors can stick with the agent after she becomes isolated. If we imagine humans with solitary *childhoods*, the objection loses its force since children in those situations have serious cognitive deficits, that is simply never true. All their choice behavior will be affected by the social practices and institutional rules that surround them. Even people who live solitary lives have at least been reared within institutional structures whose behaviors can stick with the agent in isolation. If we are to imagine humans live entirely solitary lives, the concern is remote. Our best evidence of what such an agent would be like (though how it would survive childhood is a total mystery) involves comparison with feral humans or humans who have been deliberately isolated by a captor. But these agents are known to have extreme cognitive deficits.

In sum, the STC theory explains the critical causal role that institutional rules play in determining whether observed behavior is consistent with rational choice theory. On the STC theory, individual rational choice has a primary causal role in virtue of initiating STC-economizing action; institutional rules factor into the explanation by determining the quantity of STCs in a given choice situation. STCs increase the cost of good reasoning. Consequently, STC-rational individuals become less faux-rational when these costs are high. When STCs are reduced, however, actor behavior will noticeably approach the predictions of rational choice theory. On the STC theory, institutions enable faux-rational choice by reducing STCs. It thereby rejects incorporating institutional rules into the explanation of agency. It is still we who act on our reasons and it is still we who choose whether to buy or sell some good or service. Institutions merely provide us with the relevant information by prompting us to ignore a number of bad options.

Four Problems for the STC Theory

The most powerful argument against the STC theory is that zero-intelligence traders generate efficient market outcomes. That is, they are able to make all Pareto efficient trades with almost no cognitive architecture. Given how little architecture each traditionally individuated agent contains, it seems easier to individuate actors more broadly, or at higher level of nature, such that economic agency and rationality applies to a set of processes that go beyond "the head." After all, rational choice can be generated *with no internal intelligence whatsoever*. If so, perhaps we do best to explain economic choice as flowing from an agent plus her environment rather than the agent alone. But the case of zero-intelligence traders is subject to another interpretation. Zero-intelligence traders are effective because (a) they are subject to the right set of rules and (b) they act *as if* they have completed all relevant learning. Thus, for any real agent in their same circumstances, STCs would approach zero. If STCs are near zero, then rational choice is practically free. The STC theorist can argue that by operating under these conditions, it is the zero-intelligence traders who mimic STC-rational actors, not the other way around. Consequently, we need not adopt a counterintuitive externalist view. Rationality can remain in the head even if rational decision-making is sometimes so cheap that even stupid agents can approximate it. To back up my point, Tubaro (2009) has argued that Gode and Sunder's zero-intelligence traders are constructed in ways that hide the rationality and utility increasing mechanisms at work in their simulations. In particular, Tubaro argues that the budget constraint conditions in the zero-intelligence simulations are tantamount to giving these agents some degree of rationality. So zero-intelligence traders may not be zero-intelligence after all.

Another argument against the STC theory relies on studies that show statistics courses do not usually reduce errors in probabilistic reasoning. In some cases, courses seem to increase some biases (Chiesi et al. 2009). The STC theory predicts that knowledge of statistics should reduce subjective transaction costs, and thereby increase the level of rational decision-making. But since *direct training* in statistical reasoning does not seem to aid rational decision-making, which presumably lowers STCs, then the STC theory seems to be in trouble. The defender of the STC theory

can offer three replies. First, she can argue that statistics courses do not sufficiently reduce STCs to make employing complex statistical reasoning STC-rational. Employing complex statistical reasoning will not be worthwhile if the potential gain from getting the calculation right is lower than the cognitive cost. Given that the payoffs in psychological experiments are often quite low, this may explain why subjects do not exert the cognitive effort to employ their statistical reasoning abilities. Consistent with this hypothesis, some experiments indicate that financial payoffs substantially improve rational risk-taking, although errors are not eliminated (Smith and Walker 2007) and error reduction does not increase monotonically (Gneezy and Rustichini 2000). Smith and Walker argue that "increased financial rewards [may] shift the central tendency of the data toward the predictions of rational models…[and] in virtually all cases rewards reduce the variance of the data around the predicted outcome" (Smith and Walker 2007, 245). The idea here is that if using accurate statistical reasoning increases as payments increase, then the studies are consistent with the STC theory. As the payoffs of good statistical reasoning rise, the expected utility of using one's statistical reasoning abilities associated with the deliberate, slow cognitive system increases. The STC theory predicts that faux-rational decisions should increase along the way.

The second reply involves distinguishing between two types of STCs, labor costs and capital costs. Labor costs are the costs of expending the relevant effort to make a particular calculation, whereas capital costs are the costs of improving one's calculative faculties. Some have argued that labor costs of cognition are not the only cognitive cost that matters. Camerer and Hogarth (1999) emphasize the capital aspect, which include cognitive abilities that, while largely invariant in the short run, can be improved through repetition.[30] But lower capital costs may not straightforwardly translate into lower labor costs, such that statistics courses (that arguably lower capital costs) may not improve individual rational choice. In general, the STC theory does not specify which costs are relevant, only that cognition cost, when reduced, generates faux-rational behavior. The STC theorist can also appeal to Gigerenzer and Brighton (2009) who argue that explicit reasoning by appeal to more complex forms of reasoning may not be more accurate than the use of quicker and simpler heuristics. Heuristic use may rely on more bia*s* but reduce the *variance* in the accuracy of choices in complex social situations. If explicit reasoning is not much better than heuristic use, then we have little reason to worry that increased familiarity with statistical reasoning does not increase accuracy.

A third concern about the STC theory is that it is not falsifiable. The STC theorist argues that faux-irrational behavior is STC-rational behavior. As a result, empirical evidence of irrationality can always be interpreted as faux-irrationality with high subjective transaction costs. That is, empirical evidence may always be consistent with the STC theory. For if an agent fails to maximize gains, then the STC-theorist will point out that the agent was STC-rational not to collect the information and perform the cognition necessary to realize the gain. While the agent failed to realize gains, her failure was not irrational.

[30] For an attempt to generate a ratio between these two costs, see Ortmann and Rydval (2004). They authors find around a 2:1 ratio between "capital" and "labor".

Two replies are in order. First, the STC-theory makes predictions. It suggests, for instance, that as STCs fall, rational behavior should increase. Thus, if rational choice did not increase with payoffs subject to fixed STCs, the STC theory would have a problem. Further, if lowering STCs subject to fixed payoffs did not increase faux-rational choice, then that too would spell trouble for the STC theory. We have seen just above that various forms of experimental evidence can seem inconsistent with the theory.[31]

The second reply is that appealing to varied falsifiability or verifiability constraints may not be required for good theorizing in social science. Rational choice theory is a set of principles that form a conceptual model of behavioral explanation: that is, it can potentially construe any behavior as rational depending on how we apply it to empirical data. But some applications of rational choice theory are still better than others. On analogy, recall Quine's famous *gavagai* case, where "gavagai" could refer to a rabbit or undetached rabbit parts. Observing behavior may be insufficient to decide what "gavagai" means (Quine 1970). However, we still recognize that it is more plausible to construe "gavagai" as referring to rabbits rather than undetached rabbit parts. In the same way, while we may not be able to decide between applications of rational choice theory by appealing to empirical data alone, some models will be more plausible than others.

Conclusion

Allow me to end on another note. Many readers will likely have (perhaps unjustifiably) strong opinions about the validity of much of the data and theory on the rationality of markets that I have cited. Some will be more inclined to trust the theories offered by Smith because they point to market efficiency, confirming a bias in favor of market rationality; others will be more inclined to trust the theories offered by behavioral economists like Ariely because they point to market inefficiency, confirming a bias against market rationality. In the end, one's assessment of the data may be greatly influenced by one's opinions about the general efficacy of markets and even their morality. If so, the STC theory suggests a productive method of reorienting the debate. The stereotype is that psychologists think that markets are irrational and economists think markets are rational. But on the STC theory, psychologists and economists are making a more obviously empirical, and so testable, claim: psychologists claim that STCs are high in environments where many economists think that STCs are low. Their debate can therefore be seen *not* as a debate about whether markets are rational but rather as an empirical debate about which STCs are present in a given environment. At the deepest level, they're arguing about the nature and structure of subjective transaction costs. Their debate then is not, "Are economic agents rational?" but rather, "What are the costs (STCs and otherwise) of rational decision-making?"

[31] I also believe, but I cannot elaborate here, that STCs can to some extent be measured via exertion of willpower as understood in the growing willpower literature.

Acknowledgments This paper has been long in the making, so I owe many people my thanks, including but not limited to Jerry Gaus, Terry Horgan, Keith Hankins, John Thrasher, Tyler Cowen, Adam Martin, Mike Munger, Alex Von Stein, Don Ross, Mark Alfano, Peter Boettke, Fred D'Agostino, Art Carden, David Schmidtz, Cathleen Johnson, Vernon Smith, Bart Wilson, and several anonymous referees.

References

Adams, F., & Aizawa, K. (2009). Why the mind is still in the head. In P. Robbins & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 78–95). New York: Cambridge University Press.

Alchian, A. (1950). Uncertainty, evolution and economic theory. *Journal of Political Economy, 58*, 211–221.

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York: HarperCollins.

Binmore, K. (2007). *Does game theory work? The bargaining challenge*. Cambridge: MIT Press.

Bower, B. (1996). Rational mind designs. Research into the ecology of thought treads on contested terrain. *Science News, 150*, 24–25.

Camerer, C, & Hogarth, R. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1), 7–42.

Chew, S. H., & MacCrimmon, K. R. (1979). *Alpha-Nu choice theory: A generalization of expected utility theory*. Vancouver: University of British Columbia.

Chiesi, F., Handley, S., Morsanyi, K., & Primi, C. (2009). The effects and side-effects of statistics education: Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology, 34*(3), 210–220.

Clark, A. (1996). Economic reason: The interplay of individual learning and external structure. In J. Drobak & J. Nye (Eds.), *The frontiers of the new institutional economics* (pp. 269–290). San Diego: Academic Press.

Clark, A. (1997). *Being there*. Cambridge: MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press.

Clark, A. (2012). Coupling, constitution, and the cognitive kind. In R. Menary (Ed.), *The extended mind* (pp. 81–99). Cambridge: MIT Press.

Cowen, T. (2004). How do economists think about rationality? In M. Byron (Ed.), *Satisficing and maximizing: Moral theorists on practical reason* (pp. 213–236). New York, NY: Cambridge University Press.

Denzau, A., & North, D. (1994). Shared mental models: Ideologies and institutions. *Kyklos: International Review for Social Sciences, 47*(1), 3–31.

Evans, J. (2003). In two minds: Dual-process accounts of reasoning. *Trends in the Cognitive Sciences, 7*(10), 454–459.

Friedman, M. (1953). *Essays in positive economics*. Chicago: Chiago University Press.

Gaus, G. (2007). *On philosophy, politics, and economics*. Belmont, CA: Wadsworth.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1*(1), 107–143.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650–669.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics, 115*(3), 791–810.

Gode, D., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy, 101*(1), 119–137.

Gode, D., & Sunder, S. (1997). What makes markets allocationally efficient? *The Quarterly Journal of Economics, 112*(2), 603–630.

Gould, S. J. (1992). *Bully for brontosaurus: Reflections in natural history*. Boston: W.W. Norton and Co., Inc.

Hausman, D. (1995). Rational choice and social theory: A comment. *Journal of Philosophy, 92*(2), 71–87.

Kahneman, Daniel. (2013). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.
Kahneman, D., & Tversky, A. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.
Kahneman, D., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
Kahneman, D., & Tversky, A. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.
Luce, D., & Raiffa, H. (1989). *Games and decisions: Introduction and critical survey*. Mineola, NY: Dover Publications.
Marschak, J. (1968). Economics of inquiring, communicating, deciding. *American Economic Review, 58*, 1–18.
Menary, R. (2012). Introduction: The extended mind in focus. In R. Menary (Ed.), *The extended mind* (pp. 1–26). Cambridge: MIT Press.
North, D. (2005). *Understanding the process of economic change*. Princeton: Princeton University Press.
Ortmann, A., & Rydval, O. (2004). How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Economics Letters, 85*(3), 315–320.
Polanyi, K. (2001). *The great transformation* (2nd ed., p. 1944). Boston, MA: Beacon Press.
Prinz, J. (2009). Is consciousness embodied? In R. Phillip & M. Aydede (Eds.), *The Cambridge handbook of situated cognition* (pp. 419–437). New York: Cambridge University Press.
Quine, W. V. (1970). On the reasons for indeterminacy of translation. *The Journal of Philosophy, 67*(6), 178–183.
Resnik, M. (2000). *An introduction to decision theory* (5th ed., p. 1987). Minneapolis, MN: University of Minnesota Press.
Ross, D. (2005). *Economic agency and cognitive science: Microexplanation*. Cambridge, MA: MIT Press.
Ross, D., & Ladyman, J. (2012). The alleged coupling-constitution fallacy and the mature sciences. In R. Menary (Ed.), *The extended mind* (pp. 155–166). Cambridge: MIT Press.
Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy, 101*(8), 389–428.
Sahlins, M. (1972). *Stone age economics*. Piscataway, NJ: Aldine Transaction.
Samuels, R., & Stich, S. (2004). Rationality and psychology. In A. Mele & P. Rawling (Eds.), *The oxford handbook of rationality* (pp. 279–301). New York: Oxford University Press.
Satz, D., & Ferejohn, J. (1994). Rational choice and social theory. *Journal of Philosophy, 91*(2), 71–87.
Savage, L. (1972). *The foundations of statistics*. Mineola, NY: Dover Publications.
Simon, H. (1984). *Models of bounded rationality* (Vol. I). Cambridge, MA: MIT Press.
Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3–22.
Smith, V. (1982a). Markets as economizers of information. *Economic Inquiry, 20*, 165–179.
Smith, V. (1982b). Microeconomic systems as an experimental science. *American Economic Review, 72*, 923–955.
Smith, V. (1985). Experimental economics: Reply. *American Economic Review, 75*(1), 265–272.
Smith, V. (2008). *Rationality in economics: Constructivist and ecological forms*. Cambridge: Cambridge University Press.
Smith, V., & Walker, J. (2007). Monetary rewards and decision cost in experimental economics. *Economic Inquiry, 31*(2), 245–61.
Starmer, C. (2000). Developments in non-expected utility theory. *Journal of Economic Literature, 38*, 332–382.
Thaler, R. (1994). *Quasi-rational economics*. New York: Sage.
Tubaro, P. (2009). Is individual rationality essential to market price formation? The contribution of zero-intelligence agent trading models. *Journal of Economic Methodology, 16*(1), 1–19.
Wilson, R. & Foglia, L. (2011). Embodied cognition. In E. N. Zalta (Ed.), The stanford encyclopedia of philosophy (Fall 2011 Edition). http://plato.stanford.edu/archives/fall2011/entries/embodied-cognition/.