Economics and Philosophy, **24** (2008) 381–406 Copyright © Cambridge University Press
doi:10.1017/S0266267108002034

NO REVOLUTION NECESSARY: NEURAL MECHANISMS FOR ECONOMICS

Carl F. Craver
Washington University in St. Louis

Anna Alexandrova
University of Missouri St. Louis

We argue that neuroeconomics should be a mechanistic science. We defend this view as preferable both to a revolutionary perspective, according to which classical economics is eliminated in favour of neuroeconomics, and to a classical economic perspective, according to which economics is insulated from facts about psychology and neuroscience. We argue that, like other mechanistic sciences, neuroeconomics will earn its keep to the extent that it either reconfigures how economists think about decision-making or how neuroscientists think about brain mechanisms underlying behaviour. We discuss some ways that the search for mechanisms can bring about such top-down and bottom-up revision, and we consider some examples from the recent neuroeconomics literature of how varieties of progress of this sort might be achieved.

1. INTRODUCTION[1]

Neuroeconomics is an interfield research programme. Economics and neuroscience have long been isolated from each other, with little occasion for communication across a vast cultural divide. Here we recommend a mechanistic model for bridging that divide and integrating fields in neuroeconomics.

As historians and philosophers of science, we suggest that the development of neuroscience over the last half-century (and in fact much of the history of biology) provides a model for thinking about how neuroeconomics might flourish and for thinking about how both economics and neuroscience might be changed in the process. Neuroeconomics is not the first interfield research programme. Neuroscience itself is a multifield enterprise. The different fields of neuroscience integrate their results through the effort to construct and constrain descriptions of

[1] This paper benefited immensely from comments by Erik Angner, Giacomo Bonanno, Francesco Guala, and Petri Ylkoski.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

multilevel mechanisms. Neuroeconomics can fruitfully be viewed as a continuation of that same general pattern of explanation. If one takes that perspective, then one can say in some detail what is required of an explanation in neuroeconomics, one can say how understanding lower-level mechanisms (in this case, neural mechanisms) contributes to our understanding of decision-making, and one can see how economics might potentially contribute to an understanding of neural mechanisms. Neuroeconomics could be (and perhaps should be) a mechanistic science. Our mechanistic view is closer, all things considered, to McCabe's vision than it is to Harrison's.

For the record, we make no bets on the future of neuroeconomics. The field is too young for definition let alone wagering. We do not have answers to even the most fundamental questions for defining neuroeconomics: What are its central questions? What is its domain? What are its accepted techniques? What is the appropriate vocabulary for describing phenomena in that domain? What are the standards by which research in the field will be evaluated? So it seems to us that any bets on its long-term probability of success (either in revealing hidden truths about the structure of the mind-brain or in attracting large numbers of graduate students and researchers) are not so much premature as ill defined. It is unclear what one is betting on. It is precisely because neuroeconomics is at such a young and formative stage that we hope it might benefit at this time from some perspective from the history and philosophy of neuroscience.

We are concerned in this paper exclusively with *neuroeconomics proper*, the study of the neural mechanisms of decision-making.[2] Neuroeconomics proper is the use of neuroscientific tools to study the neural mechanisms by virtue of which people (and other organisms) assess rewards, calculate future consequences, compare options, negotiate deals, perceive threats, change preferences, make choices, and so on. The goals of neuroeconomics proper are to explain economic behaviour by revealing how brain mechanisms work, how components in the brain (body, and world) work together in such a way that organisms exhibit the patterns of decision-making that they do. The term "neuroeconomics" is also used to describe *economic neural modelling*, or the export of economic concepts for use in models of brain processes or in the analysis of data delivered by neuroscientific techniques. Examples include the use of simple rational choice or Bayesian models to describe sensorimotor learning or reward circuitry (as discussed in Camerer, Loewenstein, and Prelec 2005; henceforth, CLP). Neuroeconomics proper and economic neural modelling

[2] We use "decision-making" as an umbrella term to include the evaluation of utilities and probabilities, the generation of options for action, selection among those options and among rules for decision-making. We will use the term "decision-making" in this very general way to refer to all psychological phenomena relevant to economics except where we specifically define it otherwise to discuss others' work.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

are entirely independent projects in the sense that the success of one entails nothing about the success of the other. In this paper, we recommend that neuroeconomics proper should develop as a mechanistic science. We have no commitment as to whether neural mechanisms might be illuminated through the application of the descriptive and explanatory tools of economics (as, for example, the application of equations for describing electrical circuits to neurons proved to be illuminating).

Finally, our arguments for the importance of discovering the mechanisms of decision-making do not depend on the fact that the mechanisms are neural as opposed to social, psychological, molecular, systems-level, or what have you.[3] Perhaps the mechanisms will be neural mechanisms. Perhaps Bickle (2003) is right that molecular explanations of decision-making ultimately supplant all the others (though we doubt it). Perhaps, on the other hand, the mechanisms of decision-making are most perspicuously described as psychological mechanisms or as information processing mechanisms. Much of the research discussed in the target articles for this issue and in the recent literature on neuroeconomics focuses on behavioural, not neural, economics. And when neuroscience is involved, behavioural economics is doing much of the heavy lifting. The "neural" component in these projects is in many cases limited to scanner evidence showing what areas of the brain light up when one performs some behavioural-economic task and so makes very little contribution beyond that already made by the study of decision-making behaviour. If neuroeconomics develops like the best success-stories in contemporary neuroscience (such as work on vision and on learning and memory), it will construct explanations that appeal to different levels of organization to explain different aspects of decision-making. If there is one privileged level of explanation for decision-making phenomena, the best way to find it quickly is to embrace research at multiple levels. If there are multiple levels of explanation for decision-making, the best way to find them quickly is again to do research at multiple levels. This attitude, enshrined in the mission statement of the Society for Neuroscience, helps to explain why the SfN has grown from 500 in 1969 to nearly 40 000 as of 2008.[4]

We proceed as follows. In sections 2 and 3 we argue against two visions of the relationship between neuroscience and economics: a Kuhnian

[3] If the idea of specifically cognitive or psychological mechanisms seems especially unusual, see Bechtel (2008) for a detailed account.

[4] It is also part of our multilevel perspective, though not one that we emphasize here, that empirical economists should pay attention to the markets and institutions within which individual cognitive agents act. Neuroeconomics, with its focus on brains, has an implicit bias toward internal cognitive or physiological processes in individual cognitive agents and away from the way that agents behave in institutions and in markets. In our view, a truly integrative search for mechanisms would include the study of how individuals are embedded in interpersonal, social, and otherwise environmental contexts. Such higher-level cultural and social mechanisms are susceptible to empirical investigation as well.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

revolutionary vision, in which the two disciplines are separated by a gulf of incommensurability, and a formal instrumentalist vision, according to which findings of neuroscience are irrelevant to economics. In sections 4 and 5, we present our positive vision of neuroeconomics as a mechanistic science. We explain the advantages of this vision and sketch how economics and neuroscience might gradually transform one another in the process of searching for mechanisms.

2. NO REVOLUTION NECESSARY

We reject the idea that neuroeconomics should be seen as a revolution or a "shift of paradigm" (See Glimcher 2003: 1; CLP: 55; see also Harrison's objections in this volume.) Neuroeconomists should not promise a revolution because it is too early to bet. The question of whether neuroeconomics is revolutionary is therefore not nearly as important as the question of how the science should structure itself to maximize its potential for building new and better models and for attracting and retaining converts and graduate students. Neuroeconomics should not aim to effect a revolution, we argue, because it is more promising at this stage to pursue collaborative and integrative research projects.[5]

To highlight the contrast between revolutionary and integrative research strategies, consider three of the core features of Kuhnian revolutions.[6] First, they are successional; one paradigm vanquishes and replaces another. Galileo and Copernicus did not collaborate with Aristotelians physicists and Ptolemaic astronomers to forge a new scientific perspective on the solar system; rather one global system for doing and thinking about astronomy eliminated the other (or at least that is Kuhn's narrative). Second, the rival paradigms in Kuhnian revolutions are incommensurable. They do not use the same techniques and vocabularies. They address different problems. They have different standards of evidence. They, in Kuhn's provocative phrase, inhabit different worlds. It is because the paradigms are incommensurable that revolutions must be eliminative (cf. Churchland 1989). Finally, and consequentially, the choice

[5] The term "revolution" is used in many ways to describe everything from the most momentous of scientific achievements to the most mundane changes in hair care products. Kuhn gave the term its most rigorous and provocative definition, singling out a particularly dramatic kind of scientific change. Kuhn's vision, outlined in the below three points, is our target. We intend to deny neither that neuroeconomics is potentially momentous nor that behavioural economics (leaving the "neuro" out) itself requires a fairly radical departure from business as usual for economists (see Angner and Loewenstein forthcoming). We emphasize the possibilities for communication across disparate scientific cultures rather than the barriers to such communication emphasized in Kuhn's important work. That is the crucial point here.

[6] There is room for debate as to what, exactly, Kuhn thought about each of these (see Lakatos and Musgrave 1970), and it is clear that his own thinking on the matter evolved considerably over the years (Haugland and Conant 2000). Here we need only a caricature.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

among rival paradigms cannot be grounded in any standards of evaluation mutually acceptable to members of rival paradigms. Rarely does one paradigm vanquish another by convincing the practitioners of the old to accept the new; the scientists of the old simply fail to attract students and, eventually, die. The crucial point is that revolution so conceived is in no sense a collaborative endeavour between neighbouring fields. All revolutions are victories.

Our mechanistic view of neuroeconomics differs from Kuhn's classic vision in three respects. First, we see neuroeconomics as integrative rather than successional. In this respect, a successful neuroeconomics would develop in the way Darwin and others developed the theory of evolution by natural selection, or as Watson and others worked out the genetic code. Darwin's theory integrates facts about breeding, geology, natural history, and Malthusian studies of population growth. Watson and Crick integrated diverse findings within chemistry (such as details from X-ray crystallographic findings; Chargaff's rules; basic facts about acids and bases; details about charges and chemical bonds) to elucidate the genetic code. These scientific triumphs are not victories. Second, the mechanistic model emphasizes collaboration over incommensurability. Researchers in disparate cultures build lines of communication and undertake joint projects aimed at a common collaborative goal. True incommensurability precludes collaboration; the researchers occupy different worlds and accept different standards of demonstration and evidence. Finally, in our mechanistic model, the abstract standards for evaluating causal and mechanistic explanations generally (irrespective of the details concerning any particular mechanism) are common intellectual commitments that guide the development of collaborative projects involving more or less isolated fields. The common goal of building a mechanistic model serves as the touchstone guiding the elaboration of mutually satisfactory standards for doing science and so provides the basis for communicating across the cultural divide between fields. The rhetoric of revolution emphasizes barriers to, rather than possibilities for, communication between disparate fields of science.

We sketch this mechanistic model and discuss some of its implications for the future of neuroeconomics in the final section. To see why the model is appropriate, however, it is necessary first to get clear about the goals served by economic models and, second, to see why neuroeconomists are more likely to achieve those goals if they aim at describing mechanisms.

3. GOALS OF ECONOMICS: PREDICTION, EXPLANATION, AND CONTROL

What are the goals of economics? The relevance of neuroeconomics depends on how one answers this question. And indeed economists disagree about what we should expect from economic theory. Some ways

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

of articulating the goals of economics and the status of economic theory – in particular the revealed preference view of utility (henceforth, the RPV) and the instrumentalist conception of economic theory – run counter to our vision of interfield integration. It is not that such visions of economics are intrinsically flawed. Rather they lie at the foundation of a poor strategy for the future development of economics. In this section, we explain these interpretations and give reasons to reject them.

Our argument can be summarized as follows: Suppose that the goals of economics are prediction, explanation and control. These goals are achieved better when economics aims at the discovery of mechanisms that underlie economic phenomena than when it aims merely at producing instrumental "as if" models. These mechanisms include facts about decision-making that can be and, in some cases, have been illuminated by psychology and neuroscience. Although knowledge of neural or psychological mechanisms of decision-making is not strictly necessary for providing explanations (or making good predictions, or exercising control, for that matter), the search for mechanisms is nonetheless a good strategy for building models that achieve those goals. Interpretations of economics that impede the study of mechanisms (as do, in our view, the RPV and the instrumentalist conception of economics) should not be adopted if the goal of economics is to predict, explain and control actual phenomena.

First, we explain the motivations behind the RPV and in favour of instrumentalism. Then we give reasons to think that a mechanistic interpretation is superior to the RPV if one wants to predict, explain, and control decision-making in individuals or markets. We do not claim that economic phenomena must be explained at a neural level, or any other level in particular. Rather the claim is that economics should aim at some sort of mechanistic explanation, at least in the long run, as opposed to models whose assumptions are treated as mere instruments for prediction (and derivation) and not as describing components and causal relations among them.[7]

Consider first the RPV. This view drives some of Glenn Harrison's most pressing objections to neuroeconomics. In particular, Harrison uses the RPV to argue that the discovery that people are not motivated by the search for pleasure has no bearing on economics (Harrison 2008). Here's why.[8]

[7] We do not consider the possibility of normative explanations, within which one's behaviour is explained when it is rationalized by showing that it would be reasonable for an agent to behave in such and such a way under such and such conditions. Such teleological explanations help to make behaviour intelligible at the expense of predictive accuracy (as behavioural economics appears to suggest) and the ability to control economic systems (as normative explanations fail to reveal the system's causal structure).

[8] There are many ways to interpret the RPV. One interpretation is metaphysical: preferences just are choices. A weaker interpretation is methodological: choices can and should be

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

The RPV is an interpretation of utility and preferences, the basic building blocks of models in microeconomics. Proposed in 1930s at the height of behaviourism and operationalism, the view sought to free economic theory from all reference to mental states such as preferences or utility in the sense of some psychological quantity of goodness. On the RPV, choices 'reveal' preferences – what one chooses, given certain constraints, is what one wants. In fact, as Alex Rosenberg points out, the view is a misnomer, because once preferences are identified with choices there is nothing left to reveal – talk of preferences is eliminated in favour of the talk of choices (Rosenberg 2007: 87). Of course, these choices need to satisfy certain consistency conditions – if you choose both a over b and b over a when a and b are available, then, unless you are indifferent, no respectable ranking can be constructed out of your choices. But if your choices are consistent in a particular way, then the utility function 'revealed' by them need not be interpreted as referring to a psychological quantity that motivates or explains your choices. Instead it is a formal representation of your choices.

Behaviourism and operationalism are no longer considered reasonable constraints on theories, neither in philosophy, nor in sciences of the mind. But in economics, the RPV has not gone away. One of its modern proponents argues for the RPV on the grounds that economists are ignorant of psychological mechanisms of motivation and decision making and it is best not to interpret economic theory as making any claims about those mechanisms (Binmore 2007). On the contrary, we believe this is a reason to adopt a *tentative* attitude toward economic models, but not a good reason to interpret them in a way that insulates them from empirical criticism.

How does the RPV shield economic models from criticism? It does so by claiming that models that purportedly describe relationships between utilities and expectations (or desires and beliefs), on the one hand, and choices, on the other, in fact describe relationships between choices (that is what utilities are) and other choices. So whatever a psychologist or a neuroscientist discovers about the human motivational system is irrelevant for the interpretation of economic models.

Some philosophers criticize the RPV on the grounds that it precludes the intuitive idea that mental states explain behaviour (for example, Hausman 1992; Rosenberg 2007). They also point out that the RPV cannot

used to make inferences about preferences. The methodological view can be interpreted weakly (choices are one source of evidence about preferences) or strongly (choices are the *only* source of evidence about preferences). Our targets here are the metaphysical and the strong methodological interpretations. Harrison's endorsements of Binmore and, more importantly, Samuelson make it plausible to ascribe these views to him. Whether or not Harrison or Binmore endorse the metaphysical or the strong methodological interpretations, however, we nonetheless believe that it is useful to be clear on exactly how they go wrong and, in particular, how they isolate economics from other disciplines.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

distinguish between irrational choices and changes in taste, such as when an agent's choices do not look consistent over time simply because the agent has reordered her preferences. Others argue that the RPV cannot accommodate the possibility of choices that violate one's preferences, as when one decides against one's best interest on the basis of morality or strategy (Sen 1977).

Finally, others argue that the RPV robs the economic theory of its normative force. If utility is nothing but the choices one makes, then how can we make sense of the idea that it is rational to make choices that maximize utility? Both the explanation of choices and the recommendation to make these choices rational are virtually tautologous on RPV. (See Hausman 1992 for a full discussion of these criticisms.)

We lay aside these challenges to the RPV and argue instead that the RPV cedes the explanatory ambitions of economics. We show that those who endorse the RPV face three unattractive options. They can (1) relinquish explanation as one of the goals of model-building in economics, (2) maintain that economic models are explanatory by virtue of their predictive adequacy only, or (3) maintain that economic models are explanatory according to Woodward's difference making account (Woodward 2003). The first option is transparently unsavory and, further, sacrifices the ability of economic models to serve as effective bases for intervention and control. The second option relies on a faulty view of explanation. The third option, although it relies on a compelling account of explanation, forces the supporters of the RPV to embrace shallow explanations that do not admit of empirical improvement. We think none of these options is acceptable and hence encourage economists not to use RPV to shield economics from psychology and neuroscience.

Some economists endorse the second option. Milton Friedman (1953) reconciles RPV and the explanatory value of economic models by equating explanation and prediction.[9] For Friedman, predictive adequacy is the only goal of economic theorizing. Throughout his seminal paper he uses the term "explanation" in quotation marks to indicate that explanation means prediction. For example: "Viewed as a body of substantive hypotheses, theory is to be judged by its predictive power for the class of phenomena which it is intended to 'explain'" (Friedman 1953: 184). Friedman argues that economists should build models that predict well *in the intended range of phenomena*, not that describe mechanisms or otherwise allow us to understand economic phenomena.

Friedman insists further that a model's "importance" or "significance" is inversely related to the realism of its assumptions. By importance and significance, Friedman means predictive capacity or the number of true predictions a model makes: "A hypothesis is important if it 'explains'

[9] Friedman was not trying to effect this reconciliation, but his work can be seen as doing so.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

much by little, that is, it abstracts from the common and crucial elements from the mass of complex details and permits valid predictions on the basis of them alone" (Friedman 1953: 188).

What is the intended range of phenomena? The appropriate domain of economic theory is *not* individual decision-making. Hence Friedman argues, "of course, businessmen do not actually and literally solve the system of simultaneous equations" (Friedman 1953: 193). Indeed their actions could be "habitual reaction, random chance, or whatnot" (ibid.). The relevant prediction of the maximization of returns hypothesis is not how businessmen make decisions; rather, it is their actual behaviour in the face of external circumstances. So if psychologists or neuroeconomists demonstrate that people do not have the preferences or do not use the decision-making rules that economists assume they do, that is irrelevant for evaluating economic models. Friedman thus claims that it is inappropriate for the purposes of developing economic theory to concentrate on examining and improving upon such standard assumptions of neoclassical economics as "perfect competition", "perfect monopoly", and whether businessmen do indeed reach their decisions by consulting their cost and revenue curves (Friedman 1953: 188).

Clearly, economic theory interpreted in accordance with RPV can make predictions about economic phenomena, and so, on Friedman's view, it is capable of explaining those phenomena. In his criticisms of neuroeconomics, Harrison at times embraces the idea that economic theory[10] need not be taken literally and instead should be read with an "as if" proviso. He also claims that economics need not discover mechanisms behind phenomena to be explanatory (Harrison 2008: 31). Even recognizing that sometimes standard economic theory fails to predict correctly in its intended domain of application, a Friedmanite instrumentalist rejects the need to learn about the psychological or brain mechanisms of judgment and choice. Our concern is (a) that Friedman's is not a viable view of explanation and (b) that it is not a view consistent with many of the explanations that economists routinely offer.

In the last few decades, philosophers of science have keenly debated the nature of scientific explanation. Even though there is no agreement on what explanation is, there is a lot of agreement on what it is not. When Friedman wrote his famous article, the Covering Law model of explanation dominated the philosophy of science (Hempel 1965a, 1965b). On this view, explanation is expectation; it is nothing more and nothing less than the derivation of a description of the phenomenon to be explained from premises describing laws of nature (understood as universal generalizations) and the initial and background conditions. This account is now almost universally rejected, at least as an exclusive and

[10] Friedman uses models and theories interchangeably.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

exhaustive account of explanation. To subsume a phenomenon under laws of nature is insufficient to explain it. It is universally true that all men who take birth control pills fail to get pregnant, and one could derive John's lack of pregnancy from statements describing his diligent consumption of birth control pills and this universal generalization, but the birth control pills are explanatorily irrelevant (Salmon 1971a, 1971b). It is also possible to subsume a phenomenon under laws relating effects to causes (such as the length of a shadow and the height of the flagpole), while most people believe that phenomena are explained by their causes, not their effects. The height of the flagpole explains the length of the shadow, and not vice versa (Bromberger 1966). It is also unnecessary to subsume a phenomenon under laws in order to explain it. We routinely accept explanations for why things happen while having no knowledge whatsoever about the relevant laws explaining how the things came about. We know that the knee's knocking on the table explained the inkspot on the floor even though few among us could list all of the relevant conditions and laws by virtue of which the stain came to be where it is (Scriven 1962). The ability to explain without showing that the event was to be expected on the basis of laws of nature is especially evident in probabilistic explanations (which Glimcher 2003 claims will be crucial for adequate explanations in neuroeconomics and neuroscience). Improbable things have causes and explanations though they cannot, by definition, be shown to have been expected on the basis of laws of nature (Scriven 1962; Railton 1978).

Currently the most popular alternatives to the Covering Law model all emphasize that explanations reveal causes of phenomena or the mechanisms that generate phenomena (see, e.g., Bechtel and Richardson 1993; Craver 2007; Machamer *et al*. 2000; Salmon 1984; Woodward 1989 and many others). Scientists might have an incomplete understanding of these causes and mechanisms, but to attempt to explain a phenomenon is precisely to attempt to learn what brings the phenomenon about. One variant of this view is Jim Woodward's difference-making account: roughly, x explains y if and only if had we intervened to prevent x from happening, y would not have happened (Woodward 2003). The advantage of a causal mechanical approach in general, and Woodward's interventionist version in particular, is that it unifies two of the most important goals of science – explanation and control.

Clearly, Friedman's idea that explanation is co-extensive with prediction falls prey to many of the problems that cast grave doubts on the tenability of the Covering Law model. If the Covering Law model is the only view that allows the RPV advocate to say that economics (interpreted in accordance with RPV) can provide explanations, that would be bad news since the Covering Law model fails many of the tests of genuine explanations. However, this is not right. RPV and Friedman have other options.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

Consider now the third of the above options. On this option, one allows that economic models might explain even if they do not identify the correct underlying ontology. Models involving phlogiston or gravitation might capture important explanatory truths about burning or falling without correctly describing the underlying components and activities by which things burn or fall. In like fashion, one might interpret economic models instrumentally and in accordance with RPV, but insist that such models do support the sort of counterfactuals that are necessary to make claims about how certain variables make a difference to certain others. Take, for example, a game theoretical model that claims that rational agents with privately known valuations, under conditions of certainty, plus sealed-bid first-price rules of bidding, will bid below their true valuation. On the RPV, we could read the model as making the following claim: if these agents did not have such valuations (defined as choices) they would not make such bids. However, so long as the choices that reveal preferences are not identical to the choices that are caused by these preferences, the model can be read as making a claim about how agents would behave in a variety of counterfactual circumstances: if we intervene on one set of choices, we can change another set of choices.

Indeed this result is consistent with Woodward's theory according to which to explain is to be able to answer what-if-things-had-been-different questions about the system that one is modelling. One can give such answers even if one does not know all the details of the underlying mechanisms, even if one does not couch the explanation in terms of entities at a deeper level of organization, and even if one is wrong about the ontology of that underlying explanation (Woodward 2003: 232–3). So on a difference-making account of explanation, economic models interpreted in accordance with RPV still count as explanatory, even if they are relatively shallow in the sense that they allow one to answer fewer important questions about how things would work in different circumstances than one could answer about the system being modelled (see Hitchcock and Woodward 2003). So does this mean that RPV and instrumentalist interpretations of economics can, in fact, satisfy the requirements of explanation, prediction and control?

While such models sometimes offer dim and cloudy sketches of the structure of causal dependency that must be identified in successful explanations, instrumentalism and the RPV unnecessarily insulate economics from progress in developing less dim and less cloudy models, that is, models that capture more of that causal structure more precisely. Our view makes use of Woodward's (2003) insights about the deep connection between manipulation and explanation as a crucial component in mechanistic explanation (see Craver 2007). The search for more detailed understanding of the mechanisms underlying economic phenomena potentially allows one to answer a greater range of questions

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

about what would happen if things were to be changed. Not only do mechanistic models include more measurable variables (by virtue of including variables describing the intermediate mechanisms), but the search for mechanisms is likely to call one's attention to testable differences between how-possibly models of a mechanism, and so to ever more subtle distinctions in predictions made about how people behave under a variety of circumstances. Dim and cloudy models of the causal structure underlying such behaviours, such as those that assume agents obey strict rationality assumptions and those that treat preferences as mere choices, do gesture at the causal structure of decision-making. This is why they are sometimes predictively and, to a first pass, explanatorily useful.[11] However, it is possible to be more accurate about how decision-making in fact proceeds, and an excellent way to do so is to think about the causal structures, the mechanisms, by which agents make decisions.[12]

So while we are willing to admit that Harrison's favoured interpretation of economics (in accordance with RPV and Friedman's instrumentalism) allows for some explanation, prediction, and control, we do not think that these are good interpretations if the goal is to improve the predictive and instrumental value of economics in the future (more on that record later). Understanding psychology and the brain is a better bet for such improvement. Option three, in sum, is best understood as a half step in the direction of more predictively adequate and instrumentally useful mechanistic explanations of economic phenomena.

Harrison, or another advocate of the RPV and instrumentalism, might reply that the dream of discovering causal and mechanistic explanations in economics is a philosopher's fantasy – economics in its current state is unable to live up to this impossible ideal. "As if" claims and predictions are all that economists can reasonably expect. But this reply flies in the face of the actual practice of economics.

Economists do provide successful causal explanations and do construct reliable mechanisms. The growing field of *design economics* (Roth 2002), which currently comprises the centre stage of empirical application of microeconomics, attests to this fact. When economists are called to

[11] See Mäki (1994) for a defence of something like this form of causal realism in the interpretation of economic models.

[12] Are there other types of explanation in economics? The equilibrium or arbitrage arguments is one type of explanation that does not aim at spelling out the mechanisms behind a phenomenon (e.g. Why are wages or prices the way they are? Because these are the equilibrium wages or prices.) Instead they rely on idealizing assumptions about rationality of agents and market adjustments. Whether or not these constitute a genuinely distinct type of explanation in economics, we agree with Hausman 1992 (251–252) that equilibrium explanations to be legitimate need to pay attention to the causal mechanisms that underlie equilibrium adjustments. To the extent that they don't, they are not helpful for prediction and control.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

construct reliable incentive-compatible institutions – for example, an auction that distributes spectrum licenses to telecommunication firms – they design this auction by testing a number of different causal claims. A specific kind of auction under specific conditions, which include agents' beliefs and desires, cause a specific distribution of goods and revenue generation. Crucially, in design economists' accounts of how they use theoretical models, instrumentalist "as if" interpretations are conspicuously absent (Plott 1997; Guala 2005; Alexandrova forthcoming). Such an attitude does not fit the goal of design economics to create mechanisms whose behaviour we can more or less understand and control. Granted models are treated as very tentative and merely suggestive. But they are nevertheless suggestions as to what causal factors may be relevant to one effect or another, and some of these causal factors are preferences and expectations understood psychologically, not behaviouristically. Neither models, nor the explanations that design economists construct, treat the underlying mechanisms as complete black boxes. So contrary to Harrison's claims (Harrison 2008: 31), instrumentalism is not the prevailing philosophy among practicing economists.

Perhaps design economists are mistaken; they are not entitled to any causal mechanistic claims. But if this is so, then Harrison is similarly not entitled to propose or endorse perfectly sensible causal explanations, for example, an explanation of the well-documented apparent change in people's discount rates. The explanation emphasized by behavioural and neuroeconomists is hyperbolic discounting. An alternative explanation calls into question whether experimental subjects believe the experimenters' promise to deliver the goods in the future. Harrison argues that this explanation has not been ruled out (Harrison 2008: 16–17). Maybe so, but this is not an instrumentalist attitude.

The biggest problem with formulating the goals of economics in formal instrumentalist terms is not so much philosophical as strategic. Instrumentalism isolates economics from contributions from other disciplines. A related problem arises for the current trend among neuroeconomists to appeal to Marr's three-level structure as a vision for the future of neuroeconomics. Marr's levels include a computational level, which specifies the problem to be solved, the algorithmic level, which specifies an algorithm for solving it, and an implementation level for the hardware on which the algorithm is run. Marr emphasized that each higher level is formally independent of the levels beneath it (Marr 1982). Many different algorithms can solve the same computational problem, and many different hardwares can implement the same algorithms. One can continue to investigate at the computational level independently of findings about the hardware that implements it. For this reason, in the philosophy of mind, Marr's vision has been used to argue for the irrelevance of neuroscientific evidence to the theory of mind (no matter how much Marr's practice in

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

fact violated that very assumption). In our view, it does not follow from the fact that disciplines can work independently from each other that they ought to. The value of interdisciplinary integration rather than autonomy is that it brings constraints from multiple independent perspectives to bear upon a single phenomenon and so gives it a kind of robustness or claim to reality not shared by phenomena that are not detectable from multiple independent perspectives.

Consider, for example, how Harrison uses RPV to argue that neuroeconomists' bold claims reflect a fundamental misunderstanding of the nature of economic theory, which, when corrected, undercuts the significance of neuroeconomists' claims. Findings of neuroscience and psychology, neuroeconomists emphasize, can be interpreted to suggest that the brain uses two separate though overlapping systems: one for "hedonic evaluation of stimuli" and a separate one for "motivation" (CLP: 37). If so, it is possible to be motivated to seek an object without judging that it is a greater source of pleasure than another. What bearing does this have on the central assumption of economic theory that agents choose courses of action that maximize their expected utility? Neuroeconomists claim it shows that that motivation is not grounded in pleasure seeking. However, whether or not this is relevant to economic theory depends on how utility is interpreted: interpreted as pleasure, the finding undermines the central assumption; interpreted as satisfaction of desires that may or may not be pleasurable, the finding does not bear on it; interpreted as choices, as the RPV teaches and as Harrison appears to prefer, the finding could *never* bear on it. We agree with Harrison that the finding does not *necessarily* undermine standard economic theory. But we disagree that one should choose the interpretation that precludes revision of this theory in the light of empirical findings about mechanisms that implement choice. In service of what end would one close off the possibility that one might improve the predictive or instrumental utility of one's models by studying the mechanisms by which human beings make decisions and the environmental factors to which those decision-making mechanisms do and do not respond?

It is no news that standard economic theory makes inaccurate predictions in a number of contexts. For example, game theoretical models of pretty much anything, but auctions in particular, fail to predict the bidding behaviour of real flesh and blood bidders in one-shot interactions, i.e. when the bidders are inexperienced with the interaction. The endowment effect, whereby people begin to value an object more merely because they own it, not because they derive more utility out of it, is extensively documented (Thaler 1980). So is its inconsistency with the central assumption of consumer theory that willingness to pay for an object should be equal to willingness to accept compensation if deprived of that object. Examples ranging from consumer behaviour, to

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

strategic interactions, to many other areas of economics, abound. Given the predictive failures of such standard economic models, economists cannot accept Friedman's permission to ignore the falsity of their models' assumptions. This permission is granted only to predictively successful models. How better to improve these models than through careful investigation of the psychological and brain mechanisms underlying decision-making?[13]

Consider another example. Some neuroeconomists argue that the brain is unlikely to contain a single "reasoning" module; rather, it is likely that the psychological category, "reasoning", is a dim and cloudy notion that in fact describes the behaviour of several distinct and more-or-less domain specific modules. Expected utility theory is naturally interpreted as a theory of decision-making according to which the agent first forms preferences about different outcomes represented by a cardinal utility function, then forms beliefs about probabilities of these outcomes, and then picks an act that maximizes expected utility. Psychologists argue that this account of decision-making is probably false as a descriptive theory because it fails to account for many biases, heuristics and automatic processes that underlie choice and decision-making.

Is this relevant to economic theory, which relies heavily on expected utility theory for its various models? Harrison does not think so. He gives two reasons. First, economics is about "human capital and compensating wage differentials", rather than about reasoning. And secondly, the domain-generality of reasoning is assumed merely for convenience and is not intrinsic to economics (Harrison 2008). Both reasons appeal to Friedman's idea that not all predictions of economics are on a par when it comes to testing it because not all predictions are within the intended range of the theory. We agree that sometimes it can be acceptable to ignore complexities,[14] but we urge that this should not be a general strategy of economics. The fact that some presuppositions of economics are not central to its subject matter does not change the fact that these presuppositions are problematic if the goal of the discipline is to give causal explanations and to progressively improve on the quality of those explanations.

One can grant Harrison and other critics of neuroeconomics and behavioural economics that there exists an interpretation of economic theory under which the findings of these fields (or any other fields for that matter) are irrelevant to economics and hence their significance is overblown. The question is whether this is the right interpretation. We

[13] This was Herbert Simon's reaction to Friedman's advice to ignore falsity of assumptions (Simon 1994).

[14] Indeed, we also agree that it is acceptable to build phenomenological rather than mechanistic models, and also that models that are best treated as mere instruments rather than specification of causes can be essential to many scientific projects (Craver 2006).

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

doubt it for a number of reasons: some philosophical, having to do with the nature of explanation; others strategic, having to do with how economics should best position itself in the light of the evident progress in sciences of mind and brain. The RPV and Friedman-style instrumentalism work jointly to insulate economics from possible contributions from other disciplines. In doing so, these views erect a wall of Kuhnian incommensurability between different fields. But this can only be a reasonable long-term strategy if one believed, as Gul and Pesendorfer do, that "rationality in economics is not tied to physiological causes of behavior" (Gul and Pesendorfer 2005: 24), that is, sciences of mind and brain can never bear on social sciences. But this is evidently false (how, otherwise, could advertisers be successful at influencing consumer behaviour with various psychological and physiological tricks?) or at least presumptuous. So it is best to keep economics open to contributions from other disciplines, especially sciences of mind and brain.

4. NEUROECONOMICS AS A MECHANISTIC SCIENCE

In our view, neuroeconomics should be conceived as an attempt to discover multilevel mechanisms to explain regularities in decision-making. The goal is to discover the variables (internal and external) that are relevant to explaining how people behave in different sorts of tasks and to understand how the values of those variables depend upon one another and are organized together such that they give rise to the regularities in decision-making behaviour. In neuroscience specifically, the goal is to identify the entities that are involved in these decision-making processes (the brain systems, brain regions, neurons, neurotransmitters, receptors, and so on), and to identify the activities in which those entities engage (manipulation of representations, release of neurotransmitters, binding of agonists) and how they are organized together.[15] A mechanistic neuroeconomics would proceed by searching for the mechanisms in the brain (body and world) that give rise to, sustain, modify and regulate, etc. decision-making behaviour.

Much of what is called neuroeconomics at the moment is dedicated to a very early-stage effort to sketch hypotheses about how certain decision-making mechanisms work, about which areas of the brain are involved in which decision-making tasks, and about how variations in tasks produce different patterns of brain activation. Ideally the goal is to learn which environmental variables are causally relevant to which decision-making mechanisms, and one wants to know enough about how the decision-making mechanisms work to be able to say how the mechanism will behave

[15] Here we adopt the view of mechanisms in Craver 2007. There are many others, such as Bechtel and Richardson 1993; Bunge 1997; Glennan 1996; 2002; Machamer, Darden, and Craver 2000; Woodward 2002. The differences among these are too subtle to be worth discussing in the present context.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

under a wide variety of conditions, including interventions onto the system (typically, but not always, on environmental conditions). To know how the variables depend upon one another, one needs to know how interventions to change the values of the variables in the mechanism change and do not change the values of other variables in the mechanism (see Pearl 2000; Woodward 2003). Controlled causal experiments are designed to test such relations. Our view appears to fit nicely with McCabe's idea that economists study agents' strategies while neuroscientists try to infer neural mechanisms (McCabe 2008: 6). The strategy is just the "function that the mechanisms are performing" (McCabe 2008: 7), and that can be understood as the phenomenon to be explained by the mechanism.

Mechanisms frequently span multiple levels. That is, components and activities at one level can often themselves be decomposed into organized entities and activities at a lower level, which can themselves be decomposed into organized entities and activities at still lower levels. (Relations among one set of variables can be redescribed to include more or fewer intermediate variables; the redescriptions thereby descend and ascend through levels). Researchers collaborate with one another in building such models by placing constraints on mechanisms at different levels. Researchers in one field (e.g. those who study brain systems by studying brain-damaged patients) study the gross organization of brain systems; others (such as those who use functional imaging techniques) have something to say about which locations in the brain are active during different tasks and the conditions under which those brain regions become active; others investigate neurotransmitter systems and the molecular receptors on which they act. Researchers at different levels use different techniques to place constraints on the space of plausible mechanisms for a given phenomenon. The effort to build such a multilevel model enjoins researchers to coordinate their research with work done at other levels in the same mechanisms, and this places additional constraints on any successful model. As a general sketch: the behaviour of populations might be explained in terms of the aggregation of or interaction among individual agents, whose behaviour can be explained in terms of cognitive mechanisms, which are in turn explained in terms of underlying interactions among brain regions, cells, molecules, and so on.[16] The search for multilevel mechanisms scaffolds the integration of fields.

[16] One consequence of this multilevel perspective is that neuroscience provides just one perspective on decision-making behaviour. Most individuals are embedded in social contexts, of course, and it cannot be assumed that the behaviour of populations is a simple sum of the behaviour of individual agents. The necessarily individualist perspective of neuroeconomics will need to be balanced, that is, by perspectives on how individuals work in social groups, of how they make use of environmental scaffolding (Clark 1997), and so on.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

There are three reasons neuroeconomics should be a mechanistic science. First, as noted above, the rest of neuroscience, cognitive science, and biology have adopted a largely mechanistic stance (see Darden 2006; Craver 2007; Bechtel 2008). The search for mechanisms provides a common goal toward which researchers in different fields can contribute. Anatomy, biochemistry, cytology, developmental biology, electrophysiology, evolutionary theory, molecular biology, systems neuroscience all have something to add to an understanding of mechanisms operating at multiple levels. Economics might usefully be viewed as another perspective among these,[17] adding one more kind of tile to the mosaic unity of neuroscience.

Second, the search for mechanisms is an attempt to understand the causal structures that give rise to more or less stable regularities in decision-making. While it is possible to have predictively adequate theories that do not capture accurately the causal structure of the world (Reiss 2007), generally models that allow one to infer how interventions will change the behaviour of the system will include facts about the system's causal structure: about what the relevant variables are, and about how they are causally related to one another. Mechanisms are how things work. If one knows how things work, one is potentially in a position to intervene into the system to change it for good or for ill (see Woodward 2003). In other words, if one wants instrumentally useful models, one should search for models that describe mechanisms (see Craver 2007).

Finally, even if it is possible to build predictively adequate theories that make no reference to underlying mechanisms (think for example of Snell's law, Hooke's law and Balmer's formula), one way to build more predictively adequate models is to posit underlying mechanisms. The search for mechanisms requires one to search for subtle differences in the predictions made by different models and for new experiments to bring those subtle differences out. One strategy for building more predictively adequate theories is to conceive and evaluate different models of underlying mechanisms and to test those models against the predictions of the model. This is not a failsafe strategy for generating more predictively adequate models. It might be, for example, that the hidden variables are too numerous and too varied to be included in models while still delivering models that are predictively adequate and/or manageable for predictive purposes. All we can say in response is that if there are hidden variables that are usefully included in the models or useful for clarifying when those models apply (and when they do not), then searching for mechanisms is one way to find them. It seems likely to us that there are such variables for economic theory, and so we think it would be fruitful to look for

[17] Some social scientists have argued that social science ought to proceed by discovering mechanisms (Bunge 1997; Hedström and Swedberg 1998; Hedström 2005; Elster 1989).

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

mechanisms. (To see some ways that this might be worked out in varying detail, see Woodward 2002; Craver 2007; Steel 2008.)

In short, if neuroeconomists would like to follow in the footsteps of other successful neuroscientists, if they want to build instrumentally useful models, and if they want to formulate predictively adequate models, then they have good reason to pursue a mechanistic research programme.

5. CO-EVOLUTION THROUGH INTEGRATION

While it is inappropriate to think of neuroeconomics as a revolution, we nonetheless believe that the effort to construct multilevel mechanisms in neuroeconomics will almost surely require some adjustments in the models and techniques of the parent fields if their models are to be integrated usefully. Patricia Churchland (1986) refers to this process as coevolution (see also Bechtel 1986, 1988; Craver 2007: Ch. 7). In the most radical cases, coevolution is Kuhnian elimination. But there are several varieties of coevolution short of revolution.[18] A review of recent literature in neuroeconomics shows evidence for the possibility of coevolution in neuroeconomics that is driven both from the top down (in which economics has the power to transform our neuroscientific models) and from the bottom up (in which neuroscience forces one to revise economic models).

From the top down, economics offers neuroscience fresh perspectives on the regularities in decision-making behaviour. More specifically, it offers neuroscience a library for taxonomizing economic decision-making behaviour, a set of tasks for manipulating and measuring aspects of that behaviour (from behavioural economics), and a set of representational and theoretical tools for describing and interpreting the results of such experiments. The cognitive faculties involved in these tasks are "central" faculties (such as evaluation, ranking and choice) rather than "peripheral" faculties (such as visual motion processing or control of saccadic eye movements). It is no trivial accomplishment to develop a set of tasks for measuring central cognitive phenomena precisely, a theoretical vocabulary for describing what those tasks assess, and sets of equations for describing the results of those measurements and allowing one to draw inferences from them. The tasks that economists have developed over the years to assess bargaining, price formation, cooperation, and preference reversals, for example, can be used as tasks in neuropsychological and

[18] On some holistic perspectives, any theoretical adjustment whatsoever alters the meanings of other theoretical terms and our understanding of their relations to one another and so constitutes a revolution in Kuhn's sense. We presume for the sake of argument here that there is some way to distinguish revolutions from the kind of tinkering with models constitutive of normal science.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

neuropsychiatric experiments to determine which brain mechanisms are involved in such decision-making tasks.

For example, consider how economists determine a subject's preferences. Experimental economists have developed several elicitation procedures, the most famous being the Becker–deGroot–Marschak (BDM) mechanism. This procedure asks subjects to name their reservation price for a lottery and then the lottery is auctioned off, with the proceedings going to the subject (if the auction price is higher than the subject's reservation price), or else played. BDM creates a simple task that makes it beneficial and easy for the subject to state his correct preference. Experimental economics has similar procedures for operationalizing other theoretical concepts. This *library of phenomena*, to use Francesco Guala's expression (Guala 2005), and the measurement procedures that go with it, should be a welcome addition to the neuroscientist's toolbox. Moreover, these phenomena are often described in terms of intentional categories, that is beliefs and preferences, which can function as constraints on neuroscientific explanations.[19] Clearly neuroscience potentially stands to benefit from the influx of well-characterized and quantifiable tasks for testing central decision-making mechanisms.

Viewed in the most radical light, one might see this top-down perspective as revolutionary. The thought is that economics provides a unique view of the nature of animal behaviour and so how one should break the mechanisms underlying such behaviour into components. Consider an analogy. Franz Josef Gall argued that the mind was composed of distinct faculties located in specific regions of the brain. His list of faculties is composed mostly of skills on which two individual subjects might vary independently. His faculty psychology includes artistic talent, amorousness, love for one's offspring, and other faculties that sound funny to the contemporary ear. He argued for this faculty psychology on the grounds that standard philosophical categories of mind (memory, will, cogitation) are abstractions failing to reflect the underlying modular architecture. Evolutionary psychologists argue, on like grounds, that the taxonomy of faculties implemented in the brain are likely to cross-cut the taxonomies posited by contemporary cognitive science. Economics might similarly transform how neuroscientists break the mind into nearly decomposable sub-systems and so influence what we take different regions of the brain to be doing when they perform a task. Describing the phenomenon to be explained and distinguishing it from others in a taxonomy of phenomena (such as decision-making phenomena) guide one to decompose the organism in particular ways (compare the decomposition of the body by the researcher interested in the circulatory system and the researcher interested in reaching behaviour). How one

[19] We thank Guala for making this point clear to us.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

characterizes the phenomena to be explained influences how one divides the mechanism into nearly decomposable mechanistic components (Simon 1969; Kauffman 1970). We must also acknowledge the possibility, however, that the economic perspective on human behaviour enshrined in such behavioural-economic task might not correspond in any tidy way to the mechanisms that drive human behaviour. Perhaps the brain's causal structure does not fit the economist's taxonomy just as the brain failed to accommodate Gall's faculty psychology (or so we now think). Perhaps the neural structures with which we make decisions are too diverse and too varied to be studied usefully from a neural perspective.

Between these extremes lies the possibility that economics offers a perspective on some new behavioural tasks that can be integrated into the way that cognitive scientists think about the brain and its wiring rather than requiring a radical revision of contemporary cognitive science. One might wonder, as neuroeconomists do, how decision-making is related to well-known reward systems, emotional systems, and memory systems in the brain, to name a few. Most of the neuroeconomics with which we are familiar are engaged in the effort to explore how economic phenomena are related to cognitive skills that are recognized in the current taxonomy of cognitive science. This is precisely what one would expect from a mechanistic science: the effort to show how components are organized together in such a way that they explain the phenomenon. In short, the import of economic tasks and perspectives exerts a top-down influence on neuroscience to explain how regularities in decision-making behaviour are implemented in the architecture of the brain. A prudent strategy, and one that appears to describe the practice (if not the rhetoric) of neuroeconomists, would be to try to integrate the economic understanding of behaviour into the best of ever-evolving contemporary cognitive science and to call for radical revision only when the attempted integration with what is known faces threatening anomalies.

What does neuroscience offer to drive bottom-up revision? In section 3, we argued that the search for mechanisms could lead neuroeconomists to build more predictively, explanatorily, and instrumentally useful models. This is precisely what we mean by bottom-up revision. Some types of revision involve recharacterizing the phenomenon to be explained, eliminating the phenomenon, splitting the phenomenon, lumping the phenomenon with others, and revising the scope over which a model is supposed to range.

We find very few examples of such revision in the target articles for this volume, but many examples can be found in CLP's recent review article. Again, we make no bets as to the staying power of these revisions; time will tell. Our point is that successful coevolution and revision is the benchmark of success for neuroeconomics. If neuroeconomics does not eventually produce this kind of revision, it is not worth the investment. For

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

our purposes, CLP's arguments exhibit the kind of thinking that will lead to such a productive and transformative interaction between economics and neuroscience, even if their particular examples of such coevolution do not work out.

CLP begin with a very rough taxonomy of types of processing in the brain (one which we grant for present purposes). The types differ from one another depending on whether they are controlled or automatic, on the one hand, and on whether they are cognitive or affective, on the other. Together these dimensions form four quadrants of cognitive function: controlled cognitive, controlled affective, automatic cognitive and automatic affective (CLP: 16). Many decision-making tasks, such as evaluation of utility, judgments of probability, choice, etc., straddle these quadrants. Moreover, automatic and affective dimensions, scientists are finding, influence judgment and behaviour much more than one would expect given the standard economic theory's emphasis on controlled and cognitive processes. For example, judgments about whether or not one agrees with a given editorial, how probable an event is, whether a given stimulus is positive or negative, whether consumption should be delayed or not, and many others, all take place at the intersection of several of the four quadrants (CLP). Yet only one of these processes, the controlled cognitive one, is captured by the expected utility theory.

This argument exemplifies some of the ways neuroeconomics might be expected to alter economics. First, it is an example of '*splitting*' – in which what is thought to be a unitary phenomenon is discovered to be several different phenomena explained by different mechanisms. For example, based on the distinction between automatic and controlled processes plus experimental data, neuroeconomists propose that there are two, not one, mechanisms that bring about judgments of probabilities – some explicit, driven by controlled processes, others implicit driven by automatic processes (CLP: 45). Conflict between these systems is invoked to explain psychologists' numerous observations of violation of probability axioms as, for example, in the famous Linda problem. Another possible example of splitting includes the use of different cognitive systems to explain different aspects of intertemporal choice in hyperbolic discounting (CLP: 39–40). The search for mechanisms, in short, sometimes requires one to recognize that what was previously thought to be a single phenomenon is in fact two or more phenomena.

CLP's discussion also illustrates how the search for mechanisms can lead one to *revise* the intended scope of one's model. That is, one might be forced to recognize that one's model of behaviour adequately describes the phenomenon only under a limited range of conditions. Different mechanisms are required to explain decision making outside of those conditions. Given the generality of the expected utility theory, it is tempting to view it as the laws of decision-making whenever an agent faces any

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

trade-off.[20] However, if CLP are right, only some economic decisions can accurately be described by standard economic theory. These decisions are controlled and cognitive, that is, only one of the four quadrants. Outside of that quadrant (where most decisions lie), different models are required to predict, explain and control the phenomena.

The example also illustrates how scope revision might prompt the *elimination* of certain descriptions of mechanisms. For example, a natural interpretation of some models in microeconomics is as explaining behaviour in terms of beliefs and desires (an agent does x because she desires y and believes x is a means to achieve y). But neural processes that take place in the "hedonic" quadrant (i.e. affective automatic) produce behaviour that is not easily assimilated within the standard folk psychological framework, because the affective automatic processes are cognitively and introspectively inaccessible. Indeed, people seem to confabulate reasons for their behaviour and fail to recognize factors that demonstrably influence their behaviour as having any influence on their behaviour (CLP: 31 and 38; Doris forthcoming). Of course, with elimination of incorrectly described mechanisms come proposals for better models of mechanisms. For example, racial discrimination might be seen as resulting not from conscious preferences or beliefs but from introspectively inaccessible automatic processes.

One final form of bottom-up co-evolutionary revision is lumping, the opposite of splitting. One might discover that two processes hitherto thought separate are in fact accomplished by the same or nearly the same mechanism. CLP, for example, argue that utility for money is no different neurologically than utility for other goods (CLP: 35), and that two different rules for learning in game theory, reinforcement, and learning by updating beliefs about other players, in fact employ the same neural mechanism (CLP: 50). Such a discovery can prompt researchers to see unrecognized connections between behaviours previously treated as distinct.

In short, the effort to integrate economics and neuroscience is likely to transform each in the process. We think it would be best at the moment to work to integrate, to the extent possible, the experimental approaches from economics with the best cognitive neuroscience. What counts as textbook cognitive neuroscience, especially concerning central processes, is hardly fixed in stone and exhibits considerable flexibility on even relatively short time scales. As such, one can expect that the effort to fit economics to the brain will exert downward pressure to revise how we think about the functions performed by different brain systems and regions. At the same

[20] Harrison is keen to point out that this assumption is not intrinsic to economics, but is made merely for convenience (Harrison 2008: 7). We agree, but the value of neuroeconomics does not depend on whether or not it corrects assumptions intrinsic to economics. Rather it depends on whether or not it offers better assumptions and mechanisms.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

time, the search for lower-level mechanisms forces one to ask whether one has characterized the phenomenon to be explained correctly, and so can exert bottom-up pressure to refine the higher-level taxonomy. These are the selective pressures that drive the coevolution of models and will drive the coevolution of economics and neuroscience in a mechanistic neuroeconomics.

6. CONCLUSION

We argue for a non-revolutionary approach to neuroeconomics grounded in the search for mechanisms. We argue for this perspective on the basis of the historical success of neuroscience at making progress with this model of research and on independent arguments that mechanistic models are more likely to be useful for prediction, explanation, and control than are instrumental models. We argue accordingly that economic theory should not be interpreted in such a way as to insulate it from findings in other disciplines, as the RPV would lead one to do. The search for mechanisms provides a framework for integrating results from multiple disciplines and so scaffolds the interfield connections required for coevolutionary forces to act and to transform the parent disciplines.

Of course, it does not follow from these points alone that neuroeconomics is a good idea. Indeed, some of the criticisms of the field offered by Harrison are completely independent – poor testing of hypotheses, exaggeration of new findings, misuse of brain imaging data, etc. We think that neuroeconomics should take these criticisms very seriously. The evidential and explanatory standards of neuroeconomics, which are themselves certain to evolve under interfield pressures, should be as high or higher than those of the parent disciplines at their best. Interfield projects are bound to start slowly in this respect. Researchers need to learn to speak a common language, to find mutually rewarding lines of investigation, and to develop new experimental techniques and protocols. It is also true that interfield projects often inherit the limitations of the techniques borrowed from the parent fields. These problems do not show that neuroeconomics is wrong-headed or even that neuroeconomics has so far failed to contribute to economics. These problems show instead that neuroeconomics is hard and that it will take time to learn to do it well.

REFERENCES

Angner, E. and G. Loewenstein. Forthcoming. Behavioral economics. In *Philosophy of economics*, ed. U. Mäki. Vol. 13 of Handbook of the philosophy of science, ed. D. Gabbay, P. Thagard and J. Woods. Amsterdam: Elsevier.

Alexandrova, A. Forthcoming. Making models count. *Philosophy of Science*.

Bechtel, W. 1986. The nature of cross-disciplinary research. In *Integrating scientific disciplines*, ed. W. Bechtel, 3–52. Dordrecht: Martinus Nijhoff.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

Bechtel, W. 1988. *Philosophy of science: An overview for cognitive science*. Hillsdale: NJ: Lawrence Erlbaum.
Bechtel, W. 2008. *Mental mechanisms*. Routledge.
Bechtel, W. and R. C. Richardson. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.
Bickle, J. 2003. *Philosophy and neuroscience: A ruthlessly reductive account*. Cambridge, MA: MIT Press.
Binmore, K. 2007. *Playing for real. A text on game theory.* New York: Oxford University Press.
Bromberger, S. 1966. Why-Questions. In *Readings in the philosophy of science*, ed. B. A. Brody, 66–84. Englewood Cliffs, NJ: Prentice Hall.
Bunge, M. 1997. Mechanism and explanation. *Philosophy of the Social Sciences* 27: 410–65.
Camerer, C., G. Loewenstein and D. Prelec. 2005. Neuroeconomics: how neuroscience can inform economics. *Journal of Economic Literature* XLII: 9–64.
Clark, A. 1997. *Being there: Putting mind, body, and world back together again*. Cambridge, MA: MIT Press.
Craver, C. F. 2006. When mechanistic models explain. *Synthese* 153:355–76.
Craver, C. F. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon: Oxford.
Churchland, P. M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78: 67–90.
Churchland, P. M. 1989. *A neurocomputational perspective: The nature of mind and the structure of science.* Cambridge, MA: MIT Press.
Churchland, P. S. 1986. *Neurophilosophy: Toward a unified science of the mind/.brain*. Cambridge, MA: MIT Press.
Darden, L. 2006. *Reasoning in biological discoveries: Mechanisms, interfield relations, and anomaly resolution.* New York: Cambridge University Press.
Doris, J. M. Forthcoming. Field notes: A natural history of the self. *Philosophical Issues: Metaethics*.
Elster, J. 1989. *Nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
Friedman, M. [1953] 1994. Methodology of positive economics. In *The philosophy of economics*, ed. D. M. Hausman., 145–78. Cambridge: Cambridge University Press.
Glennan, S. S. 1996. Mechanisms and the nature of causation. *Erkenntnis* 44: 49–71.
Glennan, S. S. 2002. Rethinking mechanistic explanation. *Philosophy of Science Supplement* 69: S342–53.
Glimcher, P. W. 2003. *Decisions, uncertainty and the brain: The science of neuroeconomics*. Cambridge, MA: MIT Press.
Guala, F. 2005. *Methodology of experimental economics*. Cambridge: Cambridge University Press.
Gul, F. and W. Pesendorfer. 2005. The case for mindless economics. Manuscript. Princeton, NJ: Princeton University.
Harrison, G. W. 2008. Neuroeconomics: a critical reconsideration. *Economics and Philosophy* 24.
Haugland, J. and J. Conant, eds. 2000. *The road since structure*. Chicago: Chicago University Press.
Hausman, D. M. 1992. *The inexact and separate science of economics*. Cambridge: Cambridge University Press.
Hedström, P. 2005. *Dissecting the social: On the principles of analytic sociology*. Cambridge: Cambridge University Press.
Hedström, P. and P. Swedberg. 1998. *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.
Hempel, C. 1965a. *Aspects of scientific explanation and other essays in the philosophy of science.* New York: Free Press.
Hempel, C. 1965b. Aspects of scientific explanation. In *Aspects of scientific explanation and other essays in the philosophy of science*, 331–496. New York: Free Press.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press

Hitchcock, C. and J. Woodward. 2003. Explanatory generalizations, Part 2: plumbing explanatory depth. *Nous* 37: 181–99.
Kauffman, S. A. 1970. Articulation of parts explanation in biology and the rational search for them. *Boston Studies in the Philosophy of Science* 8: 257–72.
Lakatos, I. and A. Musgrave, eds. 1970. *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.
Machamer, P. K., L. Darden and C. F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1–25.
Mäki, U. 1994. Isolation, idealization and truth in economics. In *Idealization in economics*, ed. B. Hamminga and N. de Marchi. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 38 (Special issue): 147–68.
Marr, D. 1982. *Vision*. San Francisco, CA: W.H. Freeman.
McCabe, K. 2008. Neuroeconomics and the economic sciences. *Economics and Philosophy* 24.
Pearl, J. 2000. *Causality: Models, reasoning, and inference.* Cambridge: Cambridge University Press.
Plott, C. R. 1997. Laboratory experimental testbeds: Application to the PCS auction. *Journal of Economics and Management Strategy* 6: 605–38.
Railton, P. 1978. A deductive-nomological model of probabilistic explanation. *Philosophy of Science* 45: 206–26.
Reiss, J. 2007. Do we need mechanisms in social science? *Philosophy of the Social Sciences* 37: 163–84.
Rosenberg, A. 2007. *Philosophy of social science*. Boulder, CO: Westview/Harper Collins.
Roth, A. 2002. The economist as engineer: game theory, experimental economics and computation as tools of design economics. *Econometrica* 70: 1341–78.
Salmon, W. 1971a. Statistical explanation. In *Statistical explanation and statistical relevance*, ed. W. Salmon, 29–87. Pittsburgh: University of Pittsburgh Press.
Salmon, W., ed. 1971b. *Statistical explanation and statistical relevance*. Pittsburgh: University of Pittsburgh Press.
Salmon, W. 1984. *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
Scriven, M. 1962. Explanations, predictions, and laws. In *Scientific explanation, space, and time*, ed. H. Feigl and G. Maxwell, 170–230. Vol. 3 of *Minnesota studies in the philosophy of science*. Minneapolis, MN: University of Minnesota Press.
Sen, A. 1977. Rational fools. *Philosophy and Public Affairs* 317–44.
Simon, H. 1969. *The sciences of the artificial*. Cambridge, MA: MIT Press.
Simon, H. 1994. Testability and approximation. In *The philosophy of economics*, ed. D. M. Hausman., 214–16. Cambridge: Cambridge University Press.
Steel, D. 2008. *Across the boundaries: Extrapolation in biology and social science*. Oxford: Oxford University Press.
Thaler, R. 1980. Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization* 1: 39–60.
Woodward, J. 1989. The causal/mechanical model of explanation. In *Scientific Explanation*, ed. P. Kitcher and W. Salmon. *Minnesota Studies in the Philosophy of Science* 13: 357–83.
Woodward, J. 2002. What is a mechanism? A counterfactual account. *Philosophy of Science (Supplement)* 69: S366–77.
Woodward, J. 2003. *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

https://doi.org/10.1017/S0266267108002034 Published online by Cambridge University Press