

Neuroeconomics and Confirmation Theory

(Forthcoming in *Philosophy of Science*)

Christopher Clarke *

November 1, 2013

Abstract

Neuroeconomics is a research programme founded on the thesis that cognitive and neurobiological data constitute evidence for answering economic questions. I employ confirmation theory in order to reject arguments both for and against neuroeconomics. I also emphasize that some arguments for neuroeconomics will not convince the skeptics because these arguments make a contentious assumption: economics aims for predictions and deep explanations of choices in general. I then argue for neuroeconomics by appealing to a much more restrictive (and thereby skeptic-friendly) characterization of the aims of economics.

1 Neuroeconomics and Evidence

Hypotheses about choices are often key parts of economic models. Models of the consumer for example describe how price and income determine the commodities a consumer will choose to consume. But price and income are, in a sense, *external* to the agent. This illustrates how mainstream economic models restrict themselves to these external factors. At a first approximation they do not explicitly specify how an agent's choice is determined by her *internal* states, states described either cognitively or neurobiologically.

Contrast this with McClure's (2004) model of decision-making which posits a cognitive process whereby agents compare the available options with each other; a process which takes input from two separate processes upstream.

*I am indebted to Christopher Cowie, Tim Lewens, Don Ross, and two anonymous referees for their helpful comments on this paper; and especially to Samir Okasha for his generous comments on an ancestor of this paper. I've also profited from more general discussions with Anna Alexandrova, Ken Binmore, David Papineau and Richard Pettigrew concerning the relationship between psychology and economics. This work was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement no 284123.

One of these upstream processes is modelled as fast, effortless, evolutionarily ancient and running in parallel; the other as slow, effortful, evolutionarily recent and running in series. The evolutionarily ancient process favours impulsive choices and dominates in decisions regarding immediate rewards rather than delayed rewards.

Neuroeconomics is a research programme that proposes that economists use such models of decision-making, models of how an agent's cognitive and neurobiological states determine her choices. But this proposal is controversial. In the last decade philosophers, cognitive scientists, neurobiologists and economists have created a considerable literature on the methodological issues that neuroeconomics raises. The key question under discussion is the relevance of currently available psychological data to economics. And to address this question it will be important to sharpen up the vague talk of "the relevance of psychological data to economics" that one finds in the literature.

Firstly I propose that we regard as psychological any data that go beyond standard economic "choice data". That is, data that narrow down the *specifics* of the cognitive or neurobiological mechanisms that underlie choice phenomena. So I will use "psychological data" and "cognitive and neurobiological data" interchangeably. (More on this proposal in §2 and §7.) Secondly I propose to read relevance in terms of evidential relevance. For it's uncontroversial that building cognitive or neurobiological models is relevant in that it sometimes provides economists with a creative source of inspiration. In a similar way, that is, that listening to Bob Dylan, going for a walk or watching an episode of *Lewis* might provide inspiration. The central issue—as I see it—is whether the psychological data currently used to build cognitive and neurobiological models constitutes *evidence* that can help answer the questions that mainstream economics ultimately aims to answer.

Thus some economists (Gul and Pesendorfer 2008) and philosophers (Vromen 2010a) claim that cognitive and neurobiological data provide economists with a source of creative inspiration; but that they do not constitute evidence for answering economic questions.¹ And this reflects some deeply held intuitions of practising economists (Ross 2012). Indeed it echoes the classic position advocated by Friedman and Savage (1948, 298).

This question should be contrasted with the following one: is building models of cognition or neurobiology an *efficient* means of answering economic questions? For example one might concede that psychological data

¹It is not entirely clear what Gul and Pesendorfer mean by "inspiration". Nor is it always clear how general a claim they intend to make. But for a fairly unambiguous statement of the claim see Gul and Pesendorfer (2008, 8) and Vromen (2010a, 23). What is clear is that everyone takes Gul and Pesendorfer to be denying that psychological data constitute evidence for or against economic models appropriately construed. See Ross (2011a); Ross (2011b); Harrison (2008, 322); Marchionni and Vromen (2010, 103); Dekel and Lipman (2010, 274); Bernheim (2008, 19); and Vromen (2010b, 171).

provide economics with evidence for answering economic questions, but of an inefficient sort. For standard economic data remain easier to collect than cognitive and neurobiological data. Furthermore one might think that when one has enough standard data then any cognitive and neurobiological data become redundant. Alternatively one might think that recent technological advances have made building models of cognition or neurobiology easier; so doing so is now an efficient means of answering mainstream economic questions. So economists were *historically* justified in ignoring cognition and neurobiology; but no longer so (Camerer 2008, 60).

This paper will argue in favour of neuroeconomics: currently available cognitive and neurobiological data do indeed constitute economic evidence. But I will leave the above question about efficiency unresolved, highlighting it as the locus for fruitful discussion about neuroeconomics in the future.

§2 and §3 will examine two arguments for neuroeconomics that assume an “expansive” characterization of the aims of economics. As a result I argue that these arguments will not persuade those skeptical of neuroeconomics. Instead I will propose that one accept a more restrictive characterization of the aims of economics in order to see whether neuroeconomics can still be vindicated on this restrictive characterization. The payoff will be delivered in §5 and §7, which will illustrate scenarios in which cognitive and neurobiological data do constitute economic evidence, even on a restrictive characterization, thus vindicating neuroeconomics to even its most thorough-going critics. The upshot is that neuroeconomics can be vindicated as an exercise in process-tracing (§5) and parameter estimation (§7).

This paper also makes a negative point. It will respond in §4 to an argument from neuroeconomics’ opponents, an argument whose conclusion is that a restrictive understanding of economics rules out neuroeconomics. I will employ the apparatus of confirmation theory—absent from the literature so far—to show why this argument fails. I will also use confirmation theory in §6 to undermine a common argument in *favour* of neuroeconomics.

2 Restricting the Aims of Economics

The defining/ultimate aim of chess is to checkmate one’s opponent; and the means to this end include disrupting one’s opponent’s pawn structure and capturing her pieces. Thus of any practice one can distinguish the defining/ultimate aims from the merely instrumental means to achieve these aims. So we can distinguish the question of whether it’s a defining aim of economics to describe the cognition or neurobiology of decision-making from the question of whether such descriptions of decision-making are of significant instrumental use in pursuit of economics’ aims. That is, whether cognitive and neurobiological data constitute evidence for answering economic questions,

as opposed to merely providing a source of creative inspiration.

This distinction allows one to express with greater precision an argument towards which many contributions to the literature gesture, and which takes centre stage in Aydinonat (2010, 162–65) and Craver and Alexandrova (2008, §4,5).² Argument: One of the defining aims of mainstream economics is to understand choice phenomena. That is, to predict choice events both single and multiple, particular and generic, individual and aggregate. And indeed to provide deep explanations of these events. But psychological data constitute evidence that helps one understand choice phenomena. So psychological data constitute evidence that helps economics pursue its defining aims.

The problem with the above argument is that the first premise is highly controversial. For I think most economists would on reflection deny that the defining aims of mainstream economics include the prediction or deep explanation of choice phenomena in general.³ And so a key constituency in the literature are not going to be convinced by the above argument.

It's worth pointing out why one might deny this first premise. When studying decision-making mainstream economists have described how external factors such as price and income determine an agent's choice. And the closest they have come to explicitly modelling the specifics of cognition is when they model how an agent's utility values and beliefs determine her choices.⁴ Any microeconomics textbook or any issue of the *American Economic Review* will make this clear. And so this talk of utility value and beliefs is the closest economists come to deep explanations of choices, or of predicting choices using internal factors. So mainstream economists have not attempted to specify the *series of computations* performed on internal states that eventually brings about a choice. And this is despite the fact that it has been perfectly possible to study cognition in the last fifty years using the techniques of cognitive science. The only explanation of this is that mainstream economics does not ultimately aim to specify the details of the cognitive processes of decision-making. And the same point applies to the neurobiology of decision-making. Economists therefore aim to predict and explain choices at

²Rubinstein (2003) adds the rhetorical flourish that the cognitive model he proposes seems to be more consistent with the *standard economic data* than the as-if model with which he contrasts it. See also Glimcher, Dorris, and Bayer (2005, 242); Camerer (2008, 45); Camerer (2007, C40); Bernheim (2008, 15); Gabaix and Laibson (2008, 297); and Zak (2004) for instance. Note that Aydinonat concedes that psychological data are not evidentially relevant to economic *theory* or *models*. He limits his claim to the explanation and prediction of an individual agent's choices.

³In the neuroeconomics literature see Vromen (2011, §6); Dekel and Lipman (2010, 273); and Bernheim (2008, 3) for rough agreement on this point, and for some discussion. In the economics literature see Friedman and Savage (1948) and Binmore (2009).

⁴At any rate, it's not even clear that economists' talk of utility value and beliefs is to be taken at face value as referring to internal states at all (Friedman and Savage 1948; Binmore 2009).

most via a limited class of—mostly external—factors. (See Hausman (2012) for a vigorous exposition of the opposing case.)

At any rate any such disagreement over the first premise of the above argument will in many ways be a verbal dispute over how restrictively we should apply the label “economics” or as I am putting it “mainstream economics”. The more important point is, as I’ve noted, that most economists have a restrictive understanding of economics in mind; an understanding in which economics does not ultimately aim to model cognition or neurobiology. So we can all agree that it’s an important question whether neuroeconomics can still be vindicated on this restrictive understanding of economics. Does neurobiological and cognitive data constitute evidence for economics *on the supposition that the economics of decision-making only aims ultimately to model the influence of external factors on choices?* (Alternatively: to model the influence of external factors plus that of utility value and beliefs.) In order to answer this question I’d ask Alexandrova, Aydinonat and Craver to grant for this purpose the restrictive understanding of the defining aims of economics.

But one might object as follows: “The debate over neuroeconomics concerns whether economics should aim to model the cognition and neurobiology of decision-making. But you insist on a restrictive understanding of economics. In doing so you are simply assuming that it shouldn’t model cognition and neurobiology. But why not? Isn’t it a laudable aim?” Everyone in the debate accepts that modelling cognition and neurobiology is a laudable aim that ought to be pursued. Indeed in the future this might be an aim that is pursued within university Economics departments. All that I am proposing is that we grant a particular definition of *mainstream economics at present*, a definition that doesn’t include modelling the specifics of cognition and neurobiology as a defining aim. I will have to wait until §4 to address the concern that to grant this is to automatically rule out neuroeconomics out of hand.

Now to insist upon this restrictive understanding of the defining aims of economics is to highlight a weakness in the manifestos for neuroeconomics. For these manifestos are monopolized by suggestions for how cognitive and neurobiological data might constitute evidence for or against (a) hypotheses about cognitive processes. But the authors rarely bring into focus how cognitive and neurobiological data constitute evidence for or against (b) candidate answers to the questions economics ultimately aims to answer. And on a restrictive understanding of economics (a) and (b) are very much distinct. Somehow the evidential relevance of cognitive and neurobiological data for (a) is supposed often to translate into evidential relevance for (b).⁵

⁵For example Camerer, Bhatt, and Hsu (2007, 114) says that “creating more realistic assumptions will lead to better predictions”. See Zak (2004); Dekel and Lipman (2010, 260, 264, 274); Benhabib and Bisin (2008, 321–22); Aydinonat (2010, §4); and Gabaix and

One of the reasons that this infelicity goes unnoticed I think is because of the ambiguity of the phrase “psychological data”. I have been using the phrase to refer just to cognitive and neurobiological data. This I defined in contrast to standard economic choice data, data of the form (*external factor, choice*). On my definition there is a sense in which cognitive and neurobiological data go further than choice data. For they identify—or at least further narrow down—the specifics of the cognitive or neurobiological mechanisms that underlie choice phenomena. But some authors use “psychological data” more nebulously to mean any data collected in the psychology lab, even if the data are just standard economic choice data. Take for example data about the choices agents make under lab conditions when playing a prisoner’s dilemma or an ultimatum game. And it easy to see how such “psychological data” constitute evidence relevant to (b) economic questions. For these data are just standard economic choice data. Therefore, misled by the ambiguity in “psychological data”, one might too hastily draw the conclusion that cognitive and neurobiological data constitute evidence relevant to (b) economic questions.

So, since my focus is specifically on neuroeconomics rather than on so-called behavioural economics more generally, I will define psychological data as cognitive and neurobiological data. In fact, in focusing on (b) economic questions rather than (a) cognitive hypotheses, I will be setting aside old questions about the general methodology of neuroscience. That is, I will be setting aside the long-running controversy over how data from brain scanning equipment can test *cognitive models* such as McClure’s. Instead I will be focusing on the wholly new issue raised in the neuroeconomics literature: can cognitive and psychological data answer *economic questions*?

3 As-If Construals of Economic Models

It is common to talk about testing economic models rather than to talk more broadly about answering economic questions or about the defining aims of economics. So it will be useful to say something about the content of economic models. I take it that the distinctive feature of scientific models is that their content varies for different purposes. For different purposes there are different construals of what the model says. And for the purposes of this paper the appropriate construal of an economic model will be a construal that reflects the defining aims of economics. One wants a construal such that to test an economic model is to answer a question that economics ultimately aims to answer.

Now I’ve proposed that for present purposes we grant the restrictive understanding of the defining aims of economics. So on this understanding we are to construe the model of the consumer for example as describing just

Laibson (2008) for a similar view.

the relationship between price and income and choice. (Alternatively: between price, income, utility value and choice.) Of course the model of the consumer might *appear* to say something very specific about cognitive mechanisms: agents evaluate each of the available bundles of commodities in sequence, and they keep track of the bundle that has most utility value so far in the sequence; the bundle singled out at the end of this process is the bundle the agent will choose. But for present purposes the model of the consumer is appropriately construed as not saying anything this specific about cognitive processes. In other words the appropriate construal is some form of *as if* construal, as economists typically insist (Friedman and Savage 1948). All that the model says is that the relationship between income, price (perhaps utility value) and choice is *as if* the agent performed the aforementioned computation.

The key question for neuroeconomics then becomes: can cognitive and neurobiological data constitute evidence to test models, such as the model of the consumer, *construed in the appropriate as if manner*? This observation has important implications. Consider the expected utility model in economics. One of the founding fathers of neuroeconomics, Colin Camerer, argues that one should construe this model as hypothesising that agents make decisions via a specific process, which I will label *P*. The hypothesis is that there are “two processes in the brain—one for guessing how likely one is to win and lose, and another for evaluating the hedonic pleasure and pain of winning and losing and another brain region which combines probability and hedonic sensations” (Camerer 2005, 1).

Camerer then claims we are likely to find psychological data that contradict this cognitive hypothesis. Indeed were we to find such psychological data the discovery would therefore, he claims, undermine the expected utility model in economics. The structure of Camerer’s argument is a simple *modus tollens*. An economic model entails a specific cognitive hypothesis. One then learns, using cognitive or neurobiological data, that the cognitive hypothesis is false. So one must conclude that the content of the economic model is false also.

Camerer’s argument fails for present purposes, however, and for a reason that is now obvious. For the considerations above show that for present purposes we should construe the expected utility model as just describing the relationship between external factors and choice: the relationship is *as if* economic agents compute the expected utility function via specific process *P*. (Alternatively: the relationship between external factors, utility value, beliefs and choices is as if . . .) Observe that, so construed, the expected utility model does not identify the specific computation that underlies the relationship between these factors. Thus economic models of decision-making do not entail specific cognitive hypotheses when they are construed appropri-

ately. So the first premise of Camerer's argument is false.

Despite this easy objection, arguments with Camerer's modus tollens structure are not uncommon in the literature.⁶ So a common way of reasoning in the literature is unsound. Of course to refute this way of reasoning is not to refute the more sophisticated variations upon this way of reasoning. I will wait, however, until §6 to criticise such variations upon the basic way of reasoning that Camerer exemplifies.

4 Deductivism

I've pointed out that neuroeconomic manifestos and a couple of arguments for neuroeconomics are dialectically weak: these arguments will not convince economists skeptical of neuroeconomics because the arguments implicitly assume an expansive rather than restrictive understanding of the defining aims of economics. Accordingly I've proposed that we grant the restrictive understanding of the aims of economics and then see whether neuroeconomics will be vindicated. Here's an argument that says that it won't be, an argument which I will call the deductivist argument, and which I will go on to criticise.

The Deductivist Argument. Assume—just to see what follows—the restrictive characterization of the defining aims of economics. So economic models are to be construed as describing the relationship between external environmental factors and choice. But (Deductivism) the test of an hypothesis is in the predictions it makes. More formally: a piece e of one's evidence lends some support to hypothesis h only when one is a logical consequence of the other; similarly e is evidence *against* h only when one is a logical consequence of the *negation* of the other. So e is evidence for or against an economic model only when e is a logical consequence of the environment–choice relationship the model posits (or the negation thereof or vice versa). So only data points of the form (*choice, external environmental factor*) are evidence for or against an economic model. So cognitive or neurobiological data are not evidence for or against economic models. Conclusion C : if one restricts the defining aims of economics then cognitive or neurobiological data are not evidence for or against economic models.

As many commentators note, something like this conditional C is the driving force behind Gul and Pesendorfer's (2008, 7-8) seminal criticism of neuroeconomics.⁷ And the above argument is my best attempt to reconstruct an argument for C . Ironically, some enthusiasts for neuroeconomics

⁶See Frydman et al. (2012); Glimcher, Dorris, and Bayer (2005, 219, 228); Loewenstein, Rick, and Cohen (2008, 649); and Spiegler (2008, 520) for example. Furthermore Camerer (2008, 51); Camerer (2007, C39); and Camerer, Bhatt, and Hsu (2007, 115) only make sense on the assumption that economic models entail specific cognitive hypotheses.

⁷This is noted by Harrison (2008, 322); Vromen (2011); Vromen (2010b, 172); and

also endorse the conditional C (Craver and Alexandrova 2008, §3). It's just that the enthusiasts and the critics of neuroeconomics disagree on whether the antecedent of the conditional is true, namely whether to restrict the aims of economics. And thus they can disagree over the consequent of the conditional, namely whether cognitive or neurobiological data constitute economic evidence. What I will now do is argue that both sides are too hasty in accepting this conditional. For I will reject the most obvious argument for it, the deductivist argument.

Now an epistemologist's objection to the deductivist argument, I anticipate, may be to argue that deductivism is very similar to hypothetico-deductivism; and that the literature in confirmation theory concerning the "tacking problem" has already shown that hypothetico-deductivism is false (Glymour 1980). So deductivism is false also. And so the deductivist argument fails. (Consider the case in which e confirms p . The tacking problem is that hypothetico-deductivism is committed to the following problematic thesis: e confirms $p \wedge q$, no matter what q you choose.)

But this reaction to the deductivist argument is mistaken. Hypothetico-deductivism says that e lends some support to an hypothesis p (i) when e is a logical consequence of p , and (ii) only when this is the case. Hypothetico-deductivism is therefore a much bolder thesis than deductivism. This is principally because deductivism only places a necessary condition upon evidential relevance; whereas hypothetico-deductivism also says what is sufficient for it. But it's this sufficient condition, not the necessary condition, that issues in the so-called tacking problem. So deductivism avoids the tacking problem; and our epistemologist's objection to the deductivist argument fails.

My position is to agree with this imaginary epistemologist that the deductivist argument is unsound primarily because deductivism is false. But I will provide an alternative reason to reject deductivism.

I should acknowledge my debt here to Hausman (2008, 140–42) who makes a similar point. My contribution will be to show not just *that* deductivism fails, but to invest time laying out the general epistemological reasons *why* it fails. Namely I will show that deductivism cannot handle what I will call confirmational chains. My discussion of this point will bear some structural similarity to the general discussion of confirmation theory in Laudan and Leplin (1991, 461–65); but I take my argument to be an improvement on theirs. For it avoids the decisive criticisms that Okasha (1997) levels against their argument, criticisms echoed by Psillos (1999, 164).

Consider the evidence e that market economies number one, two, three and four experience significant unemployment. And consider the inductive generalization h from the evidence: all market economies experience unem-

Crawford (2008, 249). Refer also to Gul and Pesendorfer (2008, 19–22) for potential examples.

ployment. Or rather let h be a conjunction of hypotheses $h = h_1, h_2 \dots h_n$, where for example h_5 says that market economy number five experiences unemployment. As deductivism permits, e often lends some support to h ; and h in turn lends some support to h_5 . But h_5 neither entails nor is entailed by e . Thus deductivism does not permit the support to travel through the chain, so to speak, from e to h to h_5 . That is, it does not permit e to lend any support to h_5 . But one is strongly inclined to think that it is entirely possible for e to lend support to h_5 in this case. So deductivism is false.

Indeed one can demonstrate conclusively that this is entirely possible. Imagine further that our agent suspends judgement about h and h_5 before learning e ; but after learning e she comes to accept h . Whenever a piece of evidence e lends enough support, however, to permit accepting a conjunction (h for example) it lends enough support to permit accepting each of its conjuncts (h_5 for example).⁸ It follows that e lent some support not just to h but also to h_5 . So it is entirely possible for e to lend support to h_5 in the present case.

We can run an argument of the same structure, but this time appealing to inference to the best explanation rather than enumerative induction. Take the case in which the “loveliest” explanation for a piece of one’s evidence e is a conjunction of hypotheses $h = h_1, h_2 \dots h_n$ (Lipton 1991). (Think of each conjunct as presenting a different part of a very full explanation.) As deductivism permits, e often lends some support to h ; and h in turn lends some support to h_5 . But h_5 may well neither entail nor be entailed by e . (For example h_5 might use theoretical concepts that e doesn’t use.) In this case deductivism does not permit the support to travel through the chain, so to speak, from e to h to h_5 . That is, it does not permit e to lend any support to h_5 . But I’ve already shown that this is entirely possible in this case. So again deductivism is false.

So the deductivist argument is unsound primarily because deductivism is false; and deductivism is false (for one thing) because it rules out support being passed down confirmational chains. So there is no obvious way of establishing Gul and Pesendorfer’s conditional: (C) if one restricts the defining aims of economics then cognitive or neurobiological data are not evidence for or against economic models. And this should go some distance in assuaging Craver and Alexandrova’s concerns that granting the restrictive understanding of the aims of economics will immediately settle the issue in favour of those skeptical of neuroeconomics.

One interesting way of looking at this result is as follows. The restrictive understanding of the defining aims of economics is that these aims in-

⁸I do not intend to claim that (1) whenever e lends *some* support to a conjunction, it lends some support to each of its conjuncts. Nor (2) Whenever each of two conjuncts is lent enough support to be accepted, then the conjunction is also.

volve only factors that are external to the agent rather than internal; where the boundary between internal and external is perhaps something like the boundary between inside the agent's skin and outside her skin. But Ross (2005) has argued that this version of the internal–external distinction is arbitrary: nothing of methodological importance ought to depend upon it. (Indeed he proposes his own more methodologically significant internal–external distinction.) But to argue as I have done that the restrictive understanding does not obviously rule out neuroeconomics is to argue that the “arbitrary” exclusion of factors inside the skin from the *defining* aims of economics does not obviously lead to their exclusion from economic methodology. And so my argument goes some distance in supporting Ross' claim.

5 Neuroeconomics Vindicated as Process Tracing

I've dismissed two arguments for neuroeconomics as unpersuasive, arguments that rely on an expansive understanding of the defining aims of economics. And I've just shown that assuming a more restrictive understanding of the aims of economics won't *obviously* undermine neuroeconomics. The project of this section is to assume a restrictive understanding of the aims of economics and show that neuroeconomics can be vindicated.

The basic idea can be simplified as follows. Find a cognitive or neurobiological hypothesis h that entails that this-or-that economic hypothesis is true or that such-and-such an economic hypothesis is false. (Note that this is the converse of the problematic idea discussed in §3 that economic models entail that this-or-that specific cognitive or neurobiological hypothesis is true.) Furthermore the challenge is to find such an hypothesis h that is testable using presently available neurobiological and cognitive techniques. So, put simply, establishing this cognitive or neurobiological hypothesis h would establish some economic hypotheses and undermine others. Thus neuroeconomics is vindicated. I will now illustrate exactly the kind of presently-testable cognitive or neurobiological hypothesis I have in mind, namely the hypotheses involved in process-tracing.

Consider the investor–trustee game. One agent (the investor) receives twelve pounds. And she must choose what proportion of this sum to keep and what proportion—if any—to send to another agent (the trustee). The amount sent is then tripled by the experimenter who is overseeing the game. And now the trustee must choose how this tripled amount is to be split between herself and the investor. In cases in which the investor invests a large proportion of the twelve pounds it is common to call the investor's choice “trusting”. And in cases in which the trustee acts as if she cares that the split is fair, it is common to call the trustee's preferences “other-regarding”.

Suppose that the investors' levels of oxytocin in a certain region of her

brain is strongly and positively causally relevant to her investing trustingly (Kosfeld et al. 2005). Let us further suppose that exposure to pictures of infants is strongly and positively causally relevant to oxytocin levels in this region of the brain. From this we are often licensed to accept hypothesis *C* that exposure to pictures of infants is positively causally relevant to the investor investing trustingly; at least moderately. This is process tracing in action. (Indeed I think that this process-tracing inference is guaranteed when one reads “strong causal relevance” in an “absolute” rather than an “incremental” sense; but the need for brevity prevents me from developing this undoubtedly controversial claim.)

Contrast these neurobiological suppositions with the hypothesis motivated by game-theory that says investors will choose trusting options to the extent that they have information that the trustee is other-regarding. Or rather: the hypothesis that they make such choices, give or take a small margin of error. But exposure to infant images presumably doesn’t provide an investor with information as to the trustee’s other-regardingness. So the game-theoretic hypothesis requires that exposure to pictures of infants has *negligible* causal relevance, if any, to trusting investments. And so the game-theoretic hypothesis contradicts the alternative hypothesis *C* that I’ve just shown to be supported by the neurobiological data via process tracing.

Consequently, to the extent that the neurobiological data ensure that the alternative hypothesis *C* is probable, the game-theoretic hypothesis is guaranteed to be improbable. After all, Bayesian confirmation theory tells us that whenever an hypothesis contradicts another, it is impossible for the rational degree of belief in the first and the rational degree of belief in the second to exceed one. By that same token any game-theoretic *model* whose content entails our improbable game-theoretic hypothesis will be guaranteed to be improbable. For the rational degree of belief in a proposition cannot be higher than the belief in a second proposition that the first entails.

But the above game-theoretic hypothesis, the game-theoretic model and the alternative hypothesis *C* describe just what the economics of decision-making aims to describe. For they just posit a relationship between external factors (information or exposure to images) and choices. And thus we have a scenario that illustrates how data from currently available neurobiological techniques can constitute evidence that helps economics pursue its defining aims.

Incidentally the above may be just the first step of several in the methodology employed by practicing economists. The game-theoretic hypothesis having been rendered improbable, and the alternative hypothesis having been rendered probable, one then has good reason to build a new model that incorporates this alternative hypothesis. That is, to build a model that incorporates infant images as a cause of trusting choices. And the final step will be

to test this new model, to firmly reject it or to firmly accept it. It is crucial to see that the argument of this section does not rely on saying anything about this final step, anything about what sort of data can be used to test this new model. For neuroeconomics has already been vindicated within the first step of the above methodology.

I should comment on some peculiarities of the investor-trustee scenario. Firstly I chose this scenario because it is simple, and is one of Colin Camerer's favourite examples. But Camerer describes only half of the investor-trustee scenario I described above; namely the discovery of the link between oxytocin and trusting choices. And he thinks that discovering this link is in itself sufficient to vindicate neuroeconomics. But §2 shows why this is false given a restrictive understanding of the defining aims of mainstream economics. So one needs something like the infant images part of the scenario in order to have a scenario that vindicates neuroeconomics. It is only in rare cases in the literature that anything like a full process-tracing scenario is spelt out.⁹

Secondly and consequently, the other half of the investor-trustee scenario is speculative: the link between infant images and oxytocin is pure speculation on my part for the purposes of illustration. But for our purposes this does not matter. For the infant images experiment is perfectly possible using currently available psychological techniques. So the speculative scenario still illustrates how process-tracing can allow data from currently available techniques to constitute evidence against an economic model. Finally, I do not claim that the investor-trustee scenario is an especially inspiring one; for exposure to infant images is not a very interesting factor.

I will conclude by summarizing the key features of process-tracing reasoning. One gets evidence that an external factor (exposure to infant images) is positively causally relevant to a choice factor (trust in the investor-trustee game). This is because one gets evidence that the external factor strongly is strongly and positively causally relevant to a cognitive or neurobiological factor (oxytocin levels); which in turn is strongly and positively causally relevant to the choice factor (trust in the investor-trustee game). But economics ultimately aims to describe the above relationship between external factors and choice. For example the above discovery may be inconsistent with an economic model that ignores this hitherto-neglected external factor (exposure to infant images). So this discovery is relevant to economics.

6 How not to Vindicate Neuroeconomics

The last section succeeded in vindicating neuroeconomics while assuming only a restrictive understanding of the defining aims of economics. I will now criticise another argument that tries to achieve exactly the same thing,

⁹Glimcher, Dorris, and Bayer (2005, §5.2) is a notable exception.

but fails. This argument in some ways is a more sophisticated version of the argument I attributed to Camerer in §3.

The Argument: Take again the hypothesis that expected utilities are literally computed in the brain via the specific process P envisaged in §3. Were we to learn that this specific cognitive hypothesis is true, this would lend support to the expected utility model on the appropriate construal (some form of as-if construal). For if the brain of an agent literally computes expected utility via process P then this obviously *entails* that the agent chooses as if her brain computed expected utility via process P . Imagine, however, that we instead discovered that this specific cognitive hypothesis is false. In this case the as-if expected utility model is considerably undermined. For to reject a potential support for a model is to somewhat undermine the model. So we have a scenario in which psychological data constitute evidence against an economic model.

The general idea: an economic model is considerably undermined by learning that a specific cognitive hypothesis is false; a cognitive hypothesis which, had we instead discovered to be true, would have considerably supported the model. This idea seems to me to implicitly underlie the thinking of several major neuroeconomists.¹⁰ Comparing this way of reasoning with the reasoning I criticised in §3 one sees that the advantage of the present idea is that it does not construe economic models in an inappropriately expansive rather than restrictive fashion. That is, such that they deductively entail specific cognitive hypotheses.

I will now show however that the present idea implicitly requires that we already have considerable reason to believe the following: the economic model holds just in case the specific cognitive hypothesis does. But this is not typically the case at present, construing models in the appropriate as-if manner. For example we do not at present have considerable reason to believe that the *as if* expected utility model holds just in case agents literally compute expected utility in the brain according to specific process P . After all—given our present scientific and commonsense knowledge—there are many other computational processes that are plausible as candidate mechanisms for producing expected utility behaviour.

In order to make my objection against it, I should point out that there are in fact two readings of the present idea. On the one hand we might read “the model is considerably undermined” to mean that our confidence in the model should drop to such a level $< \epsilon_0$ that warrants rejecting the model. On the other hand one might read “the model is considerably undermined” to mean that our confidence in the model should drop by a considerable degree α . I will take each reading in turn and argue that in each case the present

¹⁰This is the best way of making sense of Camerer (2008, 372); Glimcher et al. (2009, 4); and Quartz (2008, 463) for example. See also Rustichini (2005, §1).

idea implicitly requires that we already have considerable reason to believe that our economic model holds just in case our specific cognitive hypothesis does.

Applying Bayesian confirmation theory will demonstrate this for the first reading. Now the idea was that a cognitive hypothesis p would lend considerable support to our economic model M , were we to learn that the hypothesis is true. On the first reading this requires that $P(M|p) > 1 - \epsilon_1$; where $1 - \epsilon_1$ is the degree of confidence above which one is warranted in accepting M in this context.¹¹ And the idea was also that discovering this cognitive hypothesis instead to be false would considerably undermine the economic model M . On the first reading this requires that $P(M|\bar{p}) < \epsilon_0$; where ϵ_0 is the degree of confidence below which one is warranted in rejecting M in this context. But one can show that these two formal claims entail that $P(M \equiv p)$ is greater than $1 - \epsilon_0 + (\epsilon_0 - \epsilon_1)P(p)$. But ϵ_0 is low by definition, and $(\epsilon_0 - \epsilon_1)P(p)$ is low by definition irrespective of the value of $P(p)$. So $P(M \equiv p)$ will be close to one. In other words we already have considerable reason to believe that our economic model holds just in case our specific cognitive hypothesis does.

A Bayesian analysis will also demonstrate my point for the second reading. Again the idea was that a cognitive hypothesis p would lend considerable support to our economic model M , were we to learn that the hypothesis is true. On the second reading this amounts to something like $P(M|p) - P(M) = \alpha_1$ where α_1 is considerable—at least a quarter say. And the idea was also that learning instead that this cognitive hypothesis is false would considerably undermine our economic model M . On the second reading this amounts to $P(M) - P(M|\bar{p}) = \alpha_2$; where α_2 is considerable—at least a quarter say. But one can show that these two propositions together entail that

$$P(M \equiv p) = \frac{2\alpha_1\alpha_2 + \alpha_1}{\alpha_1 + \alpha_2} + P(M)\frac{\alpha_2 - \alpha_1}{\alpha_1 + \alpha_2}$$

And one can also show that in these circumstances this function is at least as large as $\frac{1}{2} + \alpha_1$ or at least as large as $\frac{1}{2} + \alpha_2$. So $P(M \equiv p)$ will be close to one. In other words we already have considerable reason to believe that our economic model holds just in case our specific cognitive hypothesis does.

So on both readings the present defence of neuroeconomics implicitly requires something that is typically false; namely that we already have considerable reason to believe that our economic model (on the appropriate as-if construal) holds just in case our specific cognitive hypothesis does. So the present defence fails.

¹¹“One should accept a proposition in a context just in case one’s degree of belief exceeds threshold ϵ .” Note that my discussion allows the threshold level to vary between contexts and propositions. To impose a fixed threshold would be highly controversial.

7 Loose Ends: Parameter Estimation

That's the main action of the paper completed. For the sake of comprehensiveness I will explore another class of scenarios in which psychological data are supposed to constitute evidence that helps economics pursue its defining aims; namely the aim of estimating the parameters of economic models. I will provide one scenario of parameter estimation that unambiguously succeeds in vindicating neuroeconomics; but I will first describe two scenarios of parameter estimation whose significance is less clear.

Scenario One. Folk tend not to spend all their savings in their old age when death is approaching. One can make sense of such behaviour by supposing that these folk prefer to bequeath something to their children. Or one can make sense of such behaviour by supposing that these folk prefer not to risk a miserable standard of living, in the case in which they live longer than expected. Are they motivated by bequests or precaution? In answer to this question one might suppose that an agent's lifelong plan—concerning when to spend and when to save—optimises a utility function with a parameter θ ; where θ denotes the extent to which that agent values making bequests to their children more than guaranteeing themselves a good standard of living.

One way of attempting to estimate an agent's θ is to ask her to introspect and report what spending–saving plan she is using. That is, experimenters might ask her to report how much she would choose to spend and to save were various contingencies to arise. Let's imagine that these introspectively-reported hypothetical choices are consistent with this agent maximizing a utility function for which θ takes value 0.3 but not for which θ takes value 0.6. So these introspective hypothetical choices suggest that θ is 0.3 for our agent. Let's call this introspection-based estimate θ^* .

Now here's the most important point: imagine that one discovers that θ^* is a good estimator of θ . (I imagine one might commission an expensive study employing a small sample of pensioners from our population. One ferrets out their detailed spending and saving records over a long time in order to measure θ . One then observes that this closely matches θ^* , the estimate based on introspective hypothetical choices. One might then extrapolate from this that θ^* is a good estimator of θ for *all* agents in our population.) And so one might estimate the distribution of θ in our population by first discovering the distribution of θ^* in our population. (To do this I imagine one might commission a cheap survey. One asks a medium-sized sample of students, workers and pensioners to answer a spending–saving questionnaire.)

But the distribution of this parameter θ in the population is the kind of question that mainstream applied economics aims to answer. For this parameter partially characterizes what options a population of agents will each choose under different external constraints e.g. under different tax policy

interventions. So we have a scenario in which introspective data constitute evidence for answering an economic question.

But do these introspective data constitute psychological data? For sure, no introspective data are standard economic data. So these data meet the first of my two necessary conditions for being psychological data (§1). But these introspective data do not seem to identify—or at least narrow down more than standard economic data does—the cognitive or neurobiological mechanisms that produce choice phenomena. All they do is purport to tell us about how an agent would *choose* under various contingencies. So they seem to fail my second necessary condition for being psychological data. What this illustrates, I think, is that the evaluation of neuroeconomics in part depends on how one understands psychological data; and the literature has left the notion very vague. So the significance of this scenario is unclear.

Scenario Two. Schotter (2008, 73–79) presents a model of decision-making in which agents interact with other agents. Firstly agents are modelled as having probabilistic beliefs about what action the other agents will engage in. Secondly agents choose the action that maximizes expected utility in light of their beliefs. Thirdly Schotter makes no further assumptions about agents' beliefs; for instance he does not assume that these beliefs are rational, as standard game theory would assume. Now Schotter estimates an agent's beliefs by asking the agent to introspect and report her beliefs—when she is suitably incentivized to tell the truth (Nyarko and Schotter 2002). (Note that these data have a better claim to be counted as psychological data because what is being introspectively reported is the agent's beliefs, not hypothetical choices as in scenario one.) And plugging these estimates of the belief parameters into Schotter's model, it turns out, leads to an economic model that is very good at predicting the agent's own choices.

And this, of course, will hold when we construe Schotter's model in an extreme *as if* manner as merely describing the choices agents will make in this interactive scenario. That is, as not describing anything to do with any internal states, not to mention cognition. It follows that there is a correlation, so to speak, between the belief parameters in this *as if* economic model and estimates of these parameters based on individuals' introspective reports. Thus we have a scenario in which (putatively) psychological data provide evidence to parameterise an economic model.

The more extreme critics of neuroeconomics might doubt, however, that mainstream economics is ultimately interested in estimating the parameters in models such as Schotter's. For these parameters describe highly transient features of agents, features that vary across short periods of time within any given agent, rather than the stable ones that one finds in scenario one. And these critics may deny that *economics* ultimately aims to measure such highly transient parameters. (Of course this is not to detract from the importance of

Schotter's work for our understanding of the psychology of decision-making.) So Schotter's work may not provide the *maximally* compelling exemplar available of the pursuit of mainstream economic aims via cognitive science. And it is to the most compelling sort of exemplars that I now turn.

Scenario Three. Consider an economic model of decision-making under uncertainty that describes an agent's choices as optimising a utility function that is not the expected utility function but rather something more complicated. The details aren't important. All that matters is that this function contains a so-called ambiguity aversion parameter. Suppose one establishes that this ambiguity aversion parameter covaries (across individual agents) with a certain neurobiological feature. For example it covaries with a pattern of activity in the agents' orbitofrontal cortices during certain choice tasks; as Hsu et al. (2005) demonstrated. As Camerer (2008, 49) suggests, it follows that if one measures the distribution of this neurobiological feature in the population, one can estimate the distribution of ambiguity aversion in that population. Thus we have a scenario in which currently available cognitive or neurobiological data constitute evidence relevant to answering a question that mainstream economics aims to answer.

Let's contrast the logical structure of the present parameter-estimation scenarios with the process-tracing scenarios of §5. In the parameter-estimation scenarios one takes it for granted that the economic model under examination is a good one. And one discovers that a psychologically-measured θ^* is a good estimator of a parameter θ in that economic model. This fact allows us to estimate the distribution of θ in the population by first estimating the distribution of θ^* . In the process-tracing scenarios, in contrast, one does not assume from the outset that the economic model is good. For example in many process tracing scenarios psychological data undermine our economic model. And they do so by pointing out that the model has ignored a crucial external cause of choice behaviour. So the process-tracing scenarios embody a bolder way in which psychological data are relevant to economics than the more modest parameter-estimation scenarios.

I should emphasize that I intend my distinction between parameter-estimation and process-tracing to shed light on the logic of confirmation underlying neuroeconomics. In actual practice neuroeconomic reasoning will typically use both sorts of reasoning in tandem. For example one might first use something similar to process-tracing to establish that a neurobiological variable is (likely to be) a good estimator of a parameter in an economic model. One might then directly confirm that this is the case, and then use this neurobiological variable for the purposes of parameter estimation.¹²

¹²Don Ross private communication.

8 Conclusion

§2 §3 and §6 rejected three arguments *for* neuroeconomics; the first two appealed to an expansive understanding of the aims of economics; the third appealed only to a restrictive understanding. And it was the use of Bayesian confirmation theory that exposed the flaw in the argument considered in §6. Similarly §4 engaged with general confirmation theoretical issues to reject an argument *against* neuroeconomics; an argument which relied on deductivism, a necessary condition on when some evidence will support an hypothesis. I showed that deductivism precludes confirmation “travelling down chains” in cases of enumerative induction and inference to the best explanation.

These are the negative claims of this paper. More positively §5 (and in a more qualified way §7) illustrated how currently available cognitive and neurobiological data can constitute evidence for answering questions that mainstream economics ultimately aims to answer: supporting and undermining economic models via process tracing (§5) and estimating the parameters of already established models (§7).

These arguments have the advantage over the arguments considered in §2 and §3, in that they do not problematically assume that economics ultimately aims to predict and deeply explain choices in general. (Instead they take it that economics ultimately aims to describe only how certain factors determine choice; and to explain choices only using these factors. These factors are external factors, with possible exception of utility value and beliefs.) The fact that these arguments do not make this problematic assumption is a good thing; because this assumption is rejected by those economists who are skeptical of neuroeconomics (§2). So the arguments of §5 and §7 will likely persuade these skeptics, unlike the arguments of §2 and §3.

References

- Aydinonat, N. Emrah. 2010. "Neuroeconomics: More than Inspiration, Less than Revolution." *Journal of Economic Methodology* 17:159–169.
- Benhabib, Jess, and Alberto Bisin. 2008. "Choice Theory and Process: Theory ahead of Measurement." In Caplin and Schotter 2008, 320–336.
- Bernheim, B. Douglas. 2008. "Neuro-economics: A Sober (but Hopeful) Appraisal." Technical Report, NBER Working Paper Series. Also later revised as Bernheim (2009).
- . 2009. "On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal." *American Economic Journal: Microeconomics* 1:1–41.
- Binmore, Ken. 2009. *Rational Decisions*. Princeton, NJ: Princeton University Press.
- Camerer, Colin F. 2005. "What is Neuroeconomics?" Unpublished Manuscript. Cited in Vromen (2010a, 34) and Gul and Pesendorfer (2008, 21).
- . 2007. "Neuroeconomics: Using Neuroscience to Make Economic Predictions." *Economic Journal* 117:C26–C42.
- . 2008. "The Case for Mindful Economics." In Caplin and Schotter 2008, 43–69.
- Camerer, Colin F, Meghana Bhatt, and Ming Hsu. 2007. "Neuroeconomics: Illustrated by the Study of Ambiguity Aversion." In *Economics and Psychology: A Promising New Cross Disciplinary Field*, ed. Bruno S. Frey and Alois Stutzer, 113–151. Cambridge, MA: MIT Press.
- Caplin, Andrew, and Andrew Schotter, eds. 2008. *The Foundations of Positive and Normative Economics: A Handbook*. Oxford: Oxford University Press.
- Craver, Carl, and Anna Alexandrova. 2008. "No Revolution Necessary: Neural Mechanisms for Economics." *Economics and Philosophy* 24:381–406.
- Crawford, Vincent P. 2008. "Look-ups as the Window of the Strategic Soul." In Caplin and Schotter 2008, 249–279.
- Dekel, Eddie, and Barton L. Lipman. 2010. "How (not) to do Decision Theory." *Annual Review of Economics* 2:257–82.
- Friedman, Milton, and L. J. Savage. 1948. "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy* 56:279–304.
- Frydman, Cary, Nicholas Barberis, Colin Camerer, Peter Bossaerts, and Antonio Rangel. 2012. "Using Neural Data to Test a Theory of Investor Behavior: An Application to Realization Utility." Technical Report, NBER Working Paper No. 18562.
- Gabaix, Xavier, and David Laibson. 2008. "The Seven Properties of Good Models." In Caplin and Schotter 2008, 292–319.

- Glimcher, Paul W., Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack. 2009. "Introduction: A Brief History of Neuroeconomics." In *Neuroeconomics: Decision Making and the Brain*, ed. Paul W. Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack, 1–12. London: Academic–Elsevier.
- Glimcher, Paul W., Michael C. Dorris, and Hannah M. Bayer. 2005. "Physiological Utility Theory and the Neuroeconomics of Choice." *Games and Economic Behavior* 52:213–256.
- Glymour, Clark. 1980. "Hypothetico-Deductivism is Hopeless." *Philosophy of Science* 47:322–325.
- Gul, Faruk, and Wolfgang Pesendorfer. 2008. "The Case for Mindless Economics." In Caplin and Schotter 2008, 3–39.
- Harrison, Glenn W. 2008. "Neuroeconomics: A Critical Reconsideration." *Economics and Philosophy* 24:303–344.
- Hausman, Daniel M. 2008. "Mindful or Mindless Economics: A Methodological Evaluation." In Caplin and Schotter 2008, 125–153.
- . 2012. *Preference, Value, Choice and Welfare*. Cambridge: Cambridge University Press.
- Hsu, Ming, Meghana Bhatt, Daniel Tranel, Ralph Adolphs, and Colin F. Camerer. 2005. "Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making." *Science* 310:1680–83.
- Kosfeld, Michael, Markus Heinrichs, Paul J. Zak, Urs Fischbacher, and Ernst Fehr. 2005. "Oxytocin Increases Trust in Humans." *Nature* 435:673–676.
- Laudan, Larry, and Jarrett Leplin. 1991. "Empirical Equivalence and Underdetermination." *Journal of Philosophy* 88:449–472.
- Lipton, Peter. 1991. *Inference to the Best Explanation*. London: Routledge.
- Loewenstein, George, Scott Rick, and Jonathan D. Cohen. 2008. "Neuroeconomics." *Annual Review of Psychology* 59:647–672.
- Marchionni, Caterina, and Jack Vromen. 2010. "Neuroeconomics: Hype or Hope?" *Journal of Economic Methodology* 17:103–106.
- McClure, Samuel M. 2004. "Separate Neural Systems Value Immediate and Delayed Reward." *Science* 306:503–07.
- Nyarko, Yaw, and Andrew Schotter. 2002. "An Experimental Study of Belief Learning Using Elicited Beliefs." *Econometrica* 70:971–1005.
- Okasha, Samir. 1997. "Laudan and Leplin on Empirical Equivalence." *British Journal for the Philosophy of Science* 48:251–256.
- Psillos, Stathis. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Quartz, Steven R. 2008. "From Cognitive Science to Cognitive Neuroscience to Neuroeconomics." *Economics and Philosophy* 24:459–471.

- Ross, Don. 2005. *Economic Theory and Cognitive Science*. Cambridge, MA: Bradford–MIT.
- . 2011a. “Cognitive Variables and Parameters in Economic Models.” In *Grounding Social Sciences in Cognitive Science*, ed. R Sun, 287–314. Cambridge, MA: MIT Press.
- . 2011b. “Neuroeconomics and Economic Methodology.” In *The Elgar Companion To Recent Economic Methodology*, ed. John B. Davis and D. Wade Hands, 61–93. Cheltenham: Edward Elgar.
- . 2012. “Mäki’s Realism and the Scope of Economics.” In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, ed. Aki Lehtinen, Jaakko Kuorikoski, and Petri Ylikoski, 181–202. Abingdon: Routledge–Taylor.
- Rubinstein, Ariel. 2003. ““Economics and Psychology”? the Case of Hyperbolic Discounting.” *International Economic Review* 44:1207–1216.
- Rustichini, Aldo. 2005. “Neuroeconomics: Present and Future.” *Games and Economic Behavior* 52:201–212.
- Schotter, Andrew. 2008. “What’s so Informative about Choice?” In Caplin and Schotter 2008, Chapter 3, 70–94.
- Spiegler, Ran. 2008. “Comments on the Potential Significance of Neuroeconomics for Economic theory.” *Economics and Philosophy*, 24:515–521.
- Vromen, Jack. 2010a. “On the Surprising Finding that Expected Utility is Literally Computed in the Brain.” *Journal of Economic Methodology* 17:17–36.
- . 2010b. “Where Economics and Neuroscience Might Meet.” *Journal of Economic Methodology* 17:171–183.
- . 2011. “Neuroeconomics: Two Camps Gradually Converging: What can Economics gain from It?” *International Review of Economics* 58:267–285.
- Zak, Paul J. 2004. “Neuroeconomics.” *Philosophical Transactions of the Royal Society London* 359:1737–1748.