Journal of Economic Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

Functionalism and the role of psychology in economics

Christopher Clarke

To cite this article: Christopher Clarke (2020): Functionalism and the role of psychology in economics, Journal of Economic Methodology, DOI: 10.1080/1350178X.2020.1798016

To link to this article: https://doi.org/10.1080/1350178X.2020.1798016

Published online: 27 Jul 2020.

Submit your article to this journal

Article views: 49

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rjec20

JOURNAL OF ECONOMIC METHODOLOGY
https://doi.org/10.1080/1350178X.2020.1798016

Routledge
Taylor & Francis Group

Check for updates

Functionalism and the role of psychology in economics

Christopher Clarke [a,b]

[a]Centre for Research in the Arts, Social Sciences, and Humanities, University of Cambridge, Cambridge, UK; [b]Erasmus School of Philosophy, Erasmus University Rotterdam, Rotterdam, Netherlands

ABSTRACT

Should economics study the psychological basis of agents' choice behaviour? I show how this question is multifaceted and profoundly ambiguous. There is no sharp distinction between 'mentalist' answers to this question and rival 'behavioural' answers. What's more, clarifying this point raises problems for mentalists of the 'functionalist' variety [Dietrich, F., & List, C. (2016). Mentalism versus behaviourism in economics: A philosophy-of-science perspective. *Economics and Philosophy*, *32*(2), 249–281. https://doi.org/10.1017/S0266267115000462]. Firstly, functionalist hypotheses collapse into hypotheses about input–output dispositions, I show, unless one places some unwelcome restrictions on what counts as a cognitive variable. Secondly, functionalist hypotheses make some risky commitments about the plasticity of agents' choice dispositions.

ARTICLE HISTORY
Received 8 April 2020
Accepted 15 July 2020

KEYWORDS
Mentalism; behaviourism; cognitive science; neuroeconomics; revealed preference; aims of economics

1. Psychological theory in economics

Having built a model that predicts the choices that one or more agents will make in a given setting, an economist might try to test her model, or to improve it, or to use her model to explain the choices that these agents make. Can psychology help economists to do this? Would it help to describe the psychological states and the psychological processes that underlie these agents' choices? The mentalist tradition of theorising about economics answers: very much so.[1] The rival behavioural tradition answers: not at all;[2] modelling choices in economics is a project that is mostly independent from the psychological study of decision-making. (It's important to note, as I do in Clarke (2016), that one can embrace the behavioural tradition without endorsing any discredited behaviourist or positivist theses.)

This controversy over the proper role of psychological concepts, explanations and evidence in economics is multifaceted and profoundly indeterminate, this paper will argue. In what way multifaceted? One facet of this controversy is the issue of whether economics should study the anatomical structures in the brain that underlie decision-making (Section 3). A second facet is the issue of whether economics should study the causes and effects of cognitive variables (Section 4). A third facet, I suggest, is the issue of whether economics should study the plasticity of agents' choice dispositions (Section 6).

In what way profoundly indeterminate? Firstly, it's not clear what one means by a cognitive variable. In particular, it's not clear what restrictions one should place on a cognitive variable counting as a genuine variable. And this indeterminacy blurs the boundaries between the mentalist and

CONTACT Christopher Clarke cjc84@cam.ac.uk; clarke@esphil.eur.nl Centre for Research in the Arts, Social Sciences, and Humanities, University of Cambridge, 7 West Road, Cambridge CB3 9DP, UK; Erasmus School of Philosophy, Erasmus University Rotterdam, Postbus 1738, 3000 DR Rotterdam, Netherlands

© 2020 Informa UK Limited, trading as Taylor & Francis Group

behavioural traditions, I will show (Section 5). Secondly, the notion of choice behaviour is ambiguous as well, in a way that blurs these boundaries even further (Section 7).

This analysis has important implications for an attractive variety of mentalism championed by Dietrich and List (2016), namely functionalism. Functionalist approaches to economic models face a dilemma, I suggest. On the one hand, if functionalists place no restrictions on what counts as a genuine cognitive variable, then functionalism collapses to an extreme form of dispositionalism, one more agreeable to the behavioural tradition (Section 5). On the other hand, if functionalists place restrictions on what counts as a genuine cognitive variable, then functionalism commits economic models to some risky and unwelcome claims about what's going on in the brain (Section 5), and to some risky claims about the plasticity of agents' choice dispositions (Section 6).

A corollary of my analysis is that the question of whether economics should study the psychological basis of choice behaviour depends – amongst other things – on the answer to five more precise questions:

(I) Does economics ultimately care about the brain anatomy that realises agents' functional–dispositional states? (Section 3)
(II) What sort of causes and effects of cognitive variables does economics ultimately care about identifying, if any? (Section 4)
(III) Does one intend to impose heavy-duty restrictions on what counts as a genuine cognitive variable? (Section 5)
(IV) Does economics ultimately care about knowing whether an agent's choice dispositions are plastic? (Section 6)
(V) Does economics ultimately aim to predict and explain an agent's choices qua intentional choices? (Section 7)

To clarify what I mean by these five questions, I will need to define the concept of the 'ultimate aims of economics' as a discipline (Section 2).

This paper builds on recent attempts within the philosophy of economics to engage with the psychology-in-economics controversy in a nuanced way. Firstly, it agrees with Guala (2019) that the dichotomy between the two traditions needs to be broken down; see also Ross (2005). Secondly, I follow Clarke (2016) and Thoma (2020) in drawing attention to an ambiguity in the concept of choice behaviour. Thirdly, I build on the work of Dietrich and List (2016) by giving an explicit example of what a functionalist hypothesis might look like in an economic context (Section 3). This example will help fix ideas, and it will help to highlight the dilemma that, I suggest, faces functionalist varieties of mentalism (Sections 5 and 6).

Fourthly, in framing the controversy as I just did, I am following Guala (2019) in interpreting the primary disagreement between the mentalist and behavioural traditions as being only secondarily about the ontology of the entities to which economic models refer. Rather, the primary disagreement is about methodology: how should economics explain choice behaviour, and what sort of evidence should it use to test such explanations? Guala's interpretation of the controversy has the advantage that it makes the controversy of direct relevance to the practice of economics.

2. The ultimate aims of economics as a discipline

Questions *I, II, IV* and *V* from my framework use the concept of 'the ultimate aims of economics as a discipline'. This section will explain what this concept means.

For the reason I gave a moment ago, I follow Guala in taking the psychology-in-economics controversy to be a controversy over what economics ought to do: should economics study the psychological basis of choice behaviour? It follows that any argument that speaks to this controversy must implicitly take the following form:

(General Argument Form)
Premise one: economics ultimately aims at *U*.
Premise two: activity *A* is a relatively good means of serving this ultimate aim *U*.
Conclusion: economics ought to perform activity *A*.
Also: performing activity *A* for reason *U* counts as doing economics.

To fix ideas, consider the following (very simple) model of choice under uncertainty.

Eve has utility 100 for outcome *A*: hearing an Aerosmith song. Eve also has utility 80 for outcome *B*: hearing a Beyonce song. Eve expects the top card in a deck of cards to be a diamond with 25 percent probability. Out of any menu of options, Eve always chooses the option with the greatest expected utility.

This economic model can be used to predict Eve's choices over two uncertain prospects. The first of these prospects is *I*: hearing Beyonce if the top card is a diamond, hearing Aerosmith otherwise. The second is *II*: hearing Aerosmith if the top card is a diamond, hearing Beyonce otherwise. Since *I* has greater expected utility than *II*, the model predicts that Eve will never choose *II* when *I* is on Eve's menu of options. Indeed, this model can also be used to predict Eve's choices over two guaranteed outcomes. The first of these guaranteed outcomes is *A*: hearing Aerosmith guaranteed. The second is *B*: hearing Beyonce guaranteed. Since *A* has greater expected utility than *I*, which has greater expected utility than *II*, which has greater expected utility than *B*, the model predicts that

Eve would never choose an option lower on the ordering *A*>*I*>*II*>*B* when an option that comes higher on this ordering is also on Eve's menu of options.

This prediction is, in effect, a mapping from menus-of-options to choices: it predicts, for each of a number of menus, the choice that Eve would make if she were to be given that menu of options. I will call this mapping Mapping One. We can now apply the General Argument Form to this model:

(Example of General Argument Form)
Premise one: with respect to this model, economics ultimately aims just to know whether Mapping One is true.
Premise two: studying the psychological basis of Eve's choice behaviour is a relatively good means of serving this ultimate aim.
Conclusion: economics ought to study the psychological basis of Eve's choice behaviour.
Also: studying the psychological basis of Eve's choice behaviour (in order to work out whether Mapping One is true) counts as doing economics.

Although some theorists will be sympathetic towards this argument (Clarke, 2014), others will reject it. For example, one might reject this argument by denying (1): perhaps economics aims to do more with this model than to raise the question of whether Mapping One is true. Or one might reject this argument by denying (2): a much better way of working out whether Mapping One is true is just to observe Eve's choices, one might contend, not to study the psychological basis of Eve's choice behaviour. Irrespective of what one thinks of this argument, however, this argument illustrates how the position one takes in the psychology-in-economics controversy is determined, in part, by what one takes the ultimate aims of economics to be. (By the by, this argument also alerts one to the possibility that psychological evidence can be used to test hypotheses – like Mapping One – that don't themselves have much psychological content.)[3]

Of course, my use of the concept of the 'ultimate aims of economics' raises the question of how to define this concept. But I'm afraid that I cannot offer a set of individually necessary and jointly sufficient conditions on something being an ultimate aim of economics. Instead, I define the concept of the ultimate aims of economics by the role that it ought to play in our reasoning about (i) what economics as a discipline ought to do, and (ii) what counts as doing economics. In particular, the concept of the ultimate aims of economics is defined by the fact that General Argument Form is a valid argument. (This is just as one cannot offer an informative set of necessary and sufficient conditions, many philosophers think, for the concept 'is a cause of' or the concept 'has a physical probability of fifty percent of occurring'. But one can nevertheless describe how one ought to reason with such concepts, thus implicitly defining them by the role that they ought to play in our reasoning.)

So, when I talk of the ultimate aims of economics, I am talking normatively and constitutively. Consider an analogy: normatively speaking, the ultimate aim of a university is to produce knowledge and to foster autonomy, one might claim. This remains the case even if, in actual fact, universities sometimes fail to live up to this aim. However, if an institution systematically falls very far short of meeting this aim, then one would not count the institution as a university. Instead, it constitutes a diploma mill, one might say. Analogously, to talk of the ultimate aims of economics is to say something normative about what economics ought to do, and something constitutive about what counts as doing economics.

What's more, just as the beliefs and desires of an institution can differ from the beliefs and desires of the members of an institution (List & Pettit, 2011), so too can institutional aims, I claim. For example, the ultimate aim of a game of football, one might claim, is that the interaction between the players is fun, and perhaps even beautiful. And this is true even if all the individual players care only about victory over the opposing team. (Indeed, in such a case it's still possible for the resulting interaction between the players to be fun and beautiful.) Analogously, the ultimate aims of economics – that is, of economics as an institution – need not be the same as the desires of economists as individuals.

Nevertheless, there must be some connection between the ultimate (normative and constitutive) aims of economics as an institution and the desires of individual economists, I'd suggest, albeit a loose one. To see this, consider the difference between the aims of economics and the aims of chemistry, say. What reason do we have for thinking that it is not an ultimate aim of economics to understand how chemicals bond with each other? It must be something to do with the institutional histories of chemistry and of economics: although some economists want to understand chemical bonding, almost none would want to use the institutional resources of an economics department to study chemical bonding, nor to publish scientific papers on bonding in an economics journal. Therefore, what individual economists do when they 'have their economics hats on' provides some evidence about the aims of economics as a discipline.

For example, this point is used by Hausman (2000, 2012) in his argument against the behavioural tradition. Hausman argues that it is very difficult to make sense of how game theorists reason game theoretically, without attributing to them the aim of describing an agent's psychology. Thus Hausman uses what individual economists do, with their economics hats on, in order to draw an inference about what economics as a discipline aims to do. (The general idea is an intuitive one (Davidson, 1973/1984). One looks at a past practice, one hypothesises about the aim that the actors might be pursuing in this practice, and then one selects the hypothetical aim that offers the most charitable way of rationalising that practice – the aim in light of which the practice appears most reasonable.)

In sum, questions *I, II, IV* and *V* from my framework are about the ultimate aims of economics. But I've argued that any argument about what economics ought to do – about what psychological concepts, explanations and evidence economics ought to be employing – must be based on some assumptions about the ultimate aims of economics. It follows that how one answers these four questions is of great important to how economics ought to be practised. I will illustrate this abstract point more concretely as I examine each question in turn.

One common way of avoiding questions such as these is to say that economics ultimately cares about understanding agents' choice behaviour, and to leave it at that. But to say this is to say very little indeed. Take, for comparison, a scientist who tells you that her discipline aims to understand why Jair Bolsonaro has COVID-19. This scientist might be an immunologist, a discipline which aims to understands how the SARS-CoV-2 virus disrupts the functioning of the lungs as a system. Or this scientist might be a microbiologist, a discipline which focuses on the cellular level, aiming to unpack the mechanism whereby the SARS-CoV-2 virus attacks cells in the human body. Or this scientist might be an epidemiologist, a discipline which aims to trace the transmission of the virus from person to person. Or the scientist might be a sociologist, a discipline which aims to explain how Bolsonaro's politics made a difference to his willingness to take precautions against contracting the virus. In other words, there is a vast network of causes of Bolsonaro getting COVID-19. No discipline aims to

identify all these causes. In general, when studying an outcome of interest, each discipline narrows its attention to a particular type of cause – its ultimate aim only being to discover which causes of that narrow type caused the outcome in question. Applying this general lesson to the case of economics and psychology: to take a position on the role of psychological concepts, explanations and data in economics, one needs to commit oneself to a position on the domain of economics. What subset of the causes of choice behaviour are the causes that economics ultimately cares about? And questions *I, II, IV, V* provide a framework of possible answers to this question.

3. Does economics care about the hardware that realises dispositions?

One way of interpreting the question 'should economics study the psychological basis of agents' choice behaviour?', I will now suggest, is to interpret it as the question 'should economics study the brain anatomy or hardware that realises an agent's functional–dispositional states?'. It follows, as a corollary, that a helpful precursor to answering the above question is to first ask question *I* from my proposed framework: does economics ultimately care about this brain anatomy or hardware? To spell out what all this means, this section will quickly rehearse some of the key ideas from the literature on functionalism in the philosophy of mind (Block, 1980; Fodor, 1987; Lewis, 1980/1983, 1994/1999; Shoemaker, 1981/1984).

The first thing I need to do is introduce the concept of a hardware variable. The following are all examples of what I will call hardware variables: the degree to which blood is flowing into the amygdala; whether or not a given neuron is firing; the type of electrical activity in the hippocampus. What makes these variables count as hardware variables is that they are defined in terms of entities (blood) interacting with other entities (the amygdala) – where these entities are themselves defined by the physical substance out of which they are made (blood contains red blood cells and white blood cells) their structure (these cells are suspended in plasma) and by their physical position in relation to other entities (the amygdala). Note also that my examples of hardware variables are all relatively local: their values are intrinsic properties of a sub-region of the brain, not of the brain as a whole. But I do not stipulate that all hardware variables must be local. Note also that my examples of hardware variables were all examples of the anatomy of an individual human. But I do not stipulate that all hardware variables must be variables relating to human anatomy. Instead, they can be variables relating to the silicone transistors in a computer, for example.

Next, I will need to introduce the concept of a hardware variable playing a causal role. To illustrate, recall the model from Section 2, and let's imagine that there are three hardware variables {U_a U_b and P} for which the following counterfactual conditionals are true.

> If variable U_a were to take any given value u_a, and variable U_b were to take any given value u_b, and variable P were to take any given value $0 < p < 1$, and if Eve's menu were any given subset of the four options *I II A B*, then Eve would choose the option with the greatest expected utility as defined by u_a u_b and p.
>
> (Means–Ends Agency Causal Role)

Of any three hardware variables, I will say that these variables play the means–ends agency causal role if and only if the counterfactual conditionals above are true of these three variables. Thus I am assuming that counterfactual conditionals, when suitably interpreted, describe causal relationships between variables (Woodward, 2003).

Similarly, for any five hardware variables {U_a, P, E, S, and D} to play the 'emotional effects causal role' they must satisfy the following counterfactual conditionals:

(Emotional Effects Causal Role)

(i) If U_a were high, then Eve would be excited ($E = 1$) if she were to hear Aerosmith. And if Eve were excited ($E = 1$), then the corners of her mouth would turn into a smile, and her heart-rate would increase.

(ii) If P were high, then Eve would be surprised ($S = 1$), if the first card in the deck were revealed not to be a Diamond. And if Eve were surprised ($S = 1$), her eyebrows would raise.

(iii) If U_a were high, and P were high, and Eve had chosen option *II*, Eve would be disappointed ($D = 1$), if she were not to hear Aerosmith. And if Eve were disappointed ($D = 1$), the corners of her mouth would droop, and her heart-rate would decrease.

Similarly, for any three hardware variables {*P*, *BS* and *TB*} to play the 'information driven causal role' they must satisfy the following counterfactual conditionals:

(Information Driven Causal Role)

(i) If Eve were shown that this deck of cards has a track record of delivering a diamond n times out of m, when shuffled, and if Eve were to believe that the deck of cards has been shuffled ($BS = 1$), then P would be roughly n/m, assuming suitably high m.

(ii) If Eve were to trust Nina's testimony ($TN = 1$), and if Nina were to tell Eve that there are n diamonds in this deck of m cards, then P would be roughly n/m.

Similarly, for any two hardware variables {U_a, U_b} to play the 'new tastes' causal role they must satisfy the following counterfactual conditionals:

(New Tastes Causal Role)

(i) If Eve had socialised in the past with people who like Aerosmith, then U_a would be higher. If Eve had socialised in the past with people who like Beyonce, then U_b would be higher.

(ii) If Aerosmith were to become more scarce a commodity, then U_a would eventually become lower. If Beyonce were to become more scarce a commodity, then U_b would eventually become lower. (See Elster (1985) on sour grapes.)

With all this in mind, one hypothesis that a behavioural scientist may ultimately aim to test is the following hypothesis:

(Hypothesis H_F) Eve is in functional–dispositional state *FD*, namely the state of being such that:

(i) there is at least one set of eight hardware variables {U_a, U_b, P, E, D, S, TN and BS} for Eve that plays these four causal roles (means–end agency, emotional effects, information driven and new tastes);

(ii) variable $U_a = 100$, $U_b = 80$, and $P = .25$.

Note that hypothesis H_F says nothing further about the physical entities whose properties these eight hardware variables denote. The location (in the brain or otherwise) of these entities remains entirely unspecified, as does the physical substance and physical structure out of which they are built. All that hypothesis H_F says is that there are eight unspecified hardware variables that play the four causal roles above. Yes, this hypothesis labels these eight variables as 'utility for Aerosmith', 'utility for Beyonce', 'probability of diamond', 'excited', 'disappointed', 'surprised', 'trusts Nina' and 'believes the deck to be shuffled'. But I included these labels purely to illustrate the above causal roles in an intuitive way. These labels can be deleted. They are not intended to add anything over and above the claim that there are eight unspecified hardware variables that play the four causal roles above. For example, 'utility for Aerosmith' is just defined in terms of these four causal roles: to say of some variable U_a that it denotes Eve's utility for Aerosmith, for example, is just to say that U_a plays these four causal roles.

In virtue of this, I will say that hypothesis H_F is a 'functionalist hypothesis', and I will say that H_F treats these eight hardware variables as 'cognitive variables', because this hypothesis picks them out merely by describing some of the causal relationships into which they enter. Also in virtue of this, I will call state *FD* itself a 'functional–dispositional state'. So 'cognitive variable' is just a label I will use for a hardware variable, when that variable is picked out causally, with the hardware features of that variable remaining unspecified. This is all that I will mean by a cognitive variable in this paper. Cognitive/hardware variables U_a U_b and P are said to 'realise' this functional–dispositional state *FD*.

It's worth pointing out that most philosophers would qualify functionalist hypothesis H_F. They would only require that these eight hardware variables tend to satisfy the counterfactuals given in these four causal roles. For example, surprise doesn't always cause raised eyebrows. I will put this complication aside in this paper.

Corollary. I can now illustrate the value of the first question in my proposed framework. Imagine that you know hypothesis H_F to be true, but you don't know the anatomy of these eight hardware variables – the location, substance, structure of whatever entities they relate to. Suppose that you then discover that variable U_a is a feature of the hippocampus. Does economics care about this discovery as an end in itself? Does this discovery about Eve speak to one of the questions about Eve that economics is ultimately aiming to answer? Camerer (2008) and Glimcher et al. (2005) would probably answer yes. Glimcher, for example, is interested as an end in itself in whether utilities are encoded in highly-localised bunches of neurons. Dietrich and List (2016) would answer no: they suggest that the only psychological hypotheses that economics ultimately cares about are functionalist hypotheses; and functionalist hypotheses abstract away from the details of the hardware, as I've just illustrated. Functionalist hypotheses treat hardware variables as cognitive variables. Economists in the behavioural tradition would clearly answer no, as would Clarke (2016), Guala (2019), Thoma (2020) and Vredenburgh (2020). For many other theorists, however, it's not clear how they would answer this question about the ultimate aims of economics.

Why is the answer to this question important? If the answer to question *I* is yes, then economics has a straightforward reason to care about evidence from the brain-scanner. EEG, fMRI and other brain scanning methods are, after all, methods designed to measure the activities of anatomical entities in the brain. If instead the answer to question *I* is no, then economics' reasons (if any) for caring about evidence from the brain scanner will be less straightforward than if the answer is yes. You might perhaps conjecture that testing a hypothesis h_1 about anatomical structures is a good means to testing some other hypothesis h_2 that economics ultimately cares about (Clarke, 2014). An example of h_{22} might be an hypothesis about how Eve will respond to various incentives. So this conjecture linking h_1 with h_2, if true, gives economics a reason to care about evidence from the brain scanner. But in this case, you will need to marshal an argument to support your conjecture linking h_1 with h_2, an argument that others may not find compelling. In contrast, if the answer is yes, then your reasons for using brain-scanner evidence are more obvious and less conjectural: economics cares about h_1 directly.

In sum, one way of interpreting the question 'should economics study the psychological basis of agents' choice behaviour?', I've suggested, is to interpret it as the question 'should economics study the brain anatomy or hardware that realises an agent's functional–dispositional states?'. And it followed, as a corollary, that a helpful precursor to answering this question is to first ask question *I* from my proposed framework: does economics ultimately care about this brain anatomy or hardware?

4. Which cognitive relationships does economics care about identifying?

An alternative way of interpreting the question 'should economics study the psychological basis of agents' choice behaviour?', I will now suggest, is to interpret it as the question 'should economics study the causes and effects of cognitive variables?'. It follows, as a corollary, that a helpful precursor to answering this question is to first ask question *II* from my proposed framework: which causes and effects of cognitive variables does economics ultimately care about identifying, if any?

Take the model of Eve from Section 2, for example. Of these three cognitive variables U_a U_b and P, one can ask: which external causes (information, socialisation) of these three variables does economics ultimately care about identifying, if any? which cognitive causes (Eve's belief that the deck has been shuffled, Eve's trust in Nina's testimony) of these three variables does economics ultimately care about identifying, if any? and which cognitive effects (excitement, disappointment, surprise) of these three variables does economics ultimately care about identifying, if any?

One answer, for example, is that economics ultimately only cares about the external causes (information, socialisation) of these three variables, but not about their cognitive causes themselves, nor about their cognitive effects. I will call this position the weakly conservative position. In this case, the weakly conservative position says that economics is ultimately interested in the new tastes role, but that it is not ultimately interested in the emotional effects role. And it says that economics is only partially interested in the information driven role; in particular, economics is not ultimately interested what the information driven role says about the cognitive variables that mediate the causal relationship between information (on the one hand) and P (on the other hand). According to the weakly conservative position, economics ultimately aims to know (i) how information makes a difference to Eve's probabilities, (ii) how factors like socialisation make a difference to her utilities, and (iii) how her probabilities and utilities and opportunities make a difference to her choices.

So, on the weakly conservative position, as far as the causal relationship between external factors and choices is concerned, economics ultimately aims to know (iv) how information, socialisation and opportunities determine Eve's choices. That is to say, to describe Eve's choices across a broad range of external circumstances – in particular a broad range of informational scenarios and a broad range of socialisation scenarios. Indeed, the functionalist hypothesis that corresponds to the weakly conservative view F_{WC} entails that, if Eve had instead been socialised with people who listen to Beyonce, then Eve would never choose an option lower on the ordering $B > II > I > A$ when an option that comes higher on this ordering is also on Eve's menu of options. This is because U_b would instead to be greater than U_a. I will call this mapping from menus-of-options to choices Mapping Two. Similarly, F_{WC} entails that, if Eve had instead information that 70 cards in 100 cards were diamonds, for example, then Eve would never choose an option lower on the ordering $A > II > I > B$ when an option that comes higher on this ordering is also on Eve's menu of options. This because P would instead be greater than 0.50. I will call this Mapping Three. (To be clear, F_{WC} 'corresponds' to the weak conservative position in the sense that F_{WC} interprets the model from Section 2 as describing only those details of Eve that the weak conservative positions says economics is ultimately interested in.)

A more strongly conservative position is that, ultimately, economics only cares about identifying the causal relationship between U_a U_b P and choice behaviour. That is to say, economics ultimately cares about the means–ends agency causal role only, rather than about any of these other causal relationships. In contrast to weak conservatism, strong conservatism entails that, as far as the causal relationship between external factors and Eve's choices are concerned, economics ultimately aims merely to describe Eve's choices in the actual informational position that she faces, and the actual socialisation scenario that she faces. It aims merely to describe her choices across variation in the opportunities that are offered to her. (That is to say, across variation in her incentives, or equivalently, the menus of options available to her.) Indeed, the functionalist hypothesis that corresponds to the strong conservative position H_{SC} entails precisely the following about Eve's choices: Eve would never choose an option lower on the ordering $A > I > II > B$ when an option that comes higher on this ordering is also on Eve's menu of options. This is Mapping One.

Imagine for example that you believe the functionalist hypothesis H_F to be true. But suppose that you then come to learn that no eight hardware variables play all four causal roles in Eve. Nevertheless, you were right that three hardware variables play the means–end agency causal role. And you were right that for these three variables $U_a = 100$, $U_b = 80$, and $P = .25$. It follows, for example, that Eve responds to incentives exactly as you expected, that is, to expansions and contractions in the opportunities (menu of options) available to her. But Eve's emotions do not respond as you imagined. And indeed her probabilities and utilities do not change over time as you expected in response to information or to socialisation. Strong conservatism entails that economics does not care about any or all of these discoveries as an end in itself. None of these discoveries about Eve speak to one of the questions about Eve that economics is ultimately aiming to answer.

It is clear that many economists take a conservative position about the aims of economics.[4] But it is not clear how strongly conservative: it seems that economics is ultimately interested in information as

an external cause of choice; but does economics ultimately care about socialisation as an external cause?

In fact, things are further complicated by the existence of an even more extreme conservative position: ultimately economics does not care about any cognitive variables at all, including U_a U_b P. So economics does not ultimately care about functionalist hypotheses such as H_F or H_{WC} or H_{SC}. Instead, economics ultimately cares only about how external factors influence choice behaviour, for example, about hypotheses such as Mapping One.

A fourth position is what I will call the radical position. The radical position is inspired by philosophical decision theorists – in particular by Lewis (1994/1999), Zynda (2000), Christensen (2004), Eriksson and Hájek (2007), and Meacham and Weisberg (2011), who argue as follows.

(1) Probabilities and utilities in philosophical decision theory are quantitative analogues of the beliefs and desires from commonsense psychology.

(2) But beliefs and desires from commonsense psychology are defined functionally by a very broad range of causal roles – much wider than the four roles I described in Section 3 – and absolutely not by the means–end agency role alone. So (3) probabilities and utilities in philosophical decision theory are defined by a very broad range of causal roles.

It follows from this argument that, insofar as economics ultimately cares about an agent's probabilities and utilities as understood by philosophical decision theory, economics cares implicitly about identifying a very broad range of causal roles. To talk about probabilities and utilities just is to talk implicitly about this broad range of causal roles, according to functionalism. Accordingly, if you think that economics ultimately cares about this very broad range of causal roles, then I will call you a radical about the aims of economics.

In passing, I want to give a reason to adopt a less-than-radical position. Namely: the causal roles to which the radical position points are ill-defined. The radical position says: whatever causal roles define the beliefs and desires of commonsense psychology, economics ultimately cares about the quantitative analogues of those causal roles for utilities and probabilities. There are two problems with this claim. The first problem is that the causal roles that define beliefs and desires are ill-defined. For example, for any given theory about the causal roles that define the variable 'believing that the MMR vaccine causes autism', there are many alternative theories that are also plausible. One theory will say that a hardware variable does not count as 'believing that the MMR vaccine causes autism' unless that variable is sensitive to testimony from medical experts; but another theory will deny this causal role is definitive of this belief. So the causal roles that define the beliefs and desires of commonsense psychology are very much contested and unclear.[5] The second problem is that for any proposed constraint on beliefs and desires, it is not clear what the quantitative analogue of that constraint is. Take for example the constraint that people choose the action that according to their beliefs will satisfy their desires. Is the quantitative analogue of this the principle of utility maximisation? Or is it instead something logically weaker such as the principle of stochastic dominance? It's unclear. So one point in favour of adopting a less-than-radical position is that doing so forces you to be clear about the causal roles that you think economics ultimately cares about identifying. (However, there are many positions that are less-than-radical without being conservative, for example, positions that allow economics to take an ultimate interest in the emotional effects role.)

Corollary. I can now illustrate the value of the second question in my proposed framework. Suppose that economics ultimately cares about identifying the emotional effects relationship. In this case, economics then has a straightforward reason to care about Eve's self report of her own phenomenology. If Eve tells you that she didn't feel very disappointed, after she expected to hear Aerosmith, but didn't, then this is evidence against an hypothesis h_1 about Eve's emotional effects, an hypothesis that economics ultimately aims to assess. In contrast, if economics doesn't care about emotional effects in themselves, then economics' reasons, if any, for caring about self-report phenomenological evidence will be less straightforward. Perhaps you might conjecture that the following is true: the fact that Eve reports that she is not disappointed here is evidence that h_2 Eve would

not choose Aerosmith in a choice between Aerosmith and Beyonce. And suppose that this hypothesis h_2 is indeed something economics does ultimately care to assess. In which case, economics does have a reason to care about self-report phenomenological evidence. However, this conjecture – that this self-report data is indeed evidence against this hypothesis h_2 – is a risky conjecture, one that others may not find compelling. In contrast, if economics does ultimately care about emotional effects in themselves, then its reasons for using self-report phenomenological evidence are obvious and much less risky.

I conclude that second way of interpreting the question 'should economics study the psychological basis of agents' choice behaviour?' is to interpret it as the question 'should economics study the cognitive causes and cognitive effects of cognitive variables?'. It follows, as a corollary, that a helpful precursor to answering this question is to first ask question *II* from my proposed framework: which cognitive causes and cognitive effects of cognitive variables does economics ultimately care about identifying, if any?

5. Does one intend to impose heavy-duty restrictions on genuine variables?

Some sets of states constitute a genuine variable and some don't. For example, the set of states {is solid, is a liquid, is a gas} constitutes the values of a genuine variable; as perhaps does the set of states {is red, is blue}. But the set of states {is a red solid or a blue liquid, is a blue solid or a red liquid} does not constitute the values of a genuine variable, many would say. In this section, I will argue that it's not clear what cognitive variables count as genuine variables. As a result, functionalist hypotheses about cognitive processes are in danger, I argue, of collapsing to hypotheses about input–output dispositions, hypotheses more agreeable to the behavioural tradition. That is to say, unless one places heavy-duty restrictions on what counts as a genuine cognitive variable. This creates a dilemma for functionalists, I suggest. And it motivates the third question in my framework: (III) does one intend to impose heavy-duty restrictions on what counts as a genuine hardware/cognitive variable?

To motivate this question, consider a functionalist hypothesis that describes how Eve's choices depend upon various external factors; for example, upon the way Eve was socialised and upon the opportunities Eve is given, that is, her menu of options. What's more, this functionalist hypothesis describes how this causal relationship is mediated by Eve's probabilities and utilities. And that's all that it describes. I will call such functionalist hypotheses 'two-step' functionalist hypotheses, because such hypotheses describe a first step from external factors to utilities and probabilities, and then a second step from utilities and probabilities to choices. (Note that functionalist hypotheses that appeal to the emotional effect role or the information driven role do not count as two-step hypotheses, because these causal roles each issue in a third step, as it were.)

Putting this point somewhat more technically, a two-step functionalist hypothesis is a long conjunction of propositions – a conjunction in which the *i*th proposition takes the following form:[6]

(Proposition α_i)

(i) If Eve were to face conjunction E_i of external factors, then genuine cognitive variable U_a would take value a_i, genuine cognitive variable U_b would take value b_i, and genuine cognitive variable P would take value p_i;

(ii) If U_a were to take value a_i, U_b were to take value b_i, and P were to take value p_i, then (regardless of the external factors E) Eve would choose C_i.

To create proposition α_1, for example, think about one possible way in which Eve might have been socialised, and one possible set of opportunities open to her. Call this conjunction of opportunities and socialisation E_1. Then fill in the choice C_1 that Eve is predicted to make in scenario E_1, as well as the value a_1 b_1 and p_1 that U_a U_b and P are predicted to have respectively in E_1. To create proposition α_2, think about another possible conjunction of opportunities and socialisation E_2 and do the same; and so on for proposition α_3 and the like.

I will now argue that two-step functionalist hypotheses are disguised hypotheses about (a) 'input–output' dispositions and also about (b) what particular hardware/cognitive variables count as genuine variables. To kick-start the argument, note that proposition α_i clearly entails:

(Proposition β_i) There is some hardware state of Eve's H_i, such that:

(i) Necessarily hardware state H_i occurs if and only if genuine cognitive variable U_a takes value a_i and genuine cognitive variable U_b takes value b_i and genuine cognitive variable P takes value p_i;
(ii) If Eve were to face conjunction E_i of external factors, then she would be in hardware state H_i;
(iii) If Eve were in hardware state H_i, then (regardless of the external factors E) she would choose C_i.

But let $\mathrm{E}[U_a = a_i]$ denote the set of those (conjunctions of) external factors that each cause $U_a = a_i$, according to the functionalist hypothesis in question. For example, if the functionalist hypothesis says that E_1 causes $U_a = a_i$, then E_1 will be included in set $\mathrm{E}[U_a = a_i]$. What's more, let $\mathrm{H}[U_a = a_i]$ denote the set of hardware states that are each caused – as a matter of fact – by at least one of the (conjunctions of) external factors in $\mathrm{E}[U_a = a_i]$. For example, if the functionalist hypothesis says that E_1 causes $U_a = a_i$, and if as a matter of fact E_1 causes H_1, then H_1 will be included in set $\mathrm{H}[U_a = a_i]$. Note that it follows from these definitions that, if the functionalist hypothesis in question is true, genuine cognitive variable U_a takes the value a_i if and only if one of the hardware states in set $\mathrm{H}[U_a = a_i]$ occurs. The point of defining things in this way is that the claim that 'a state in $\mathrm{H}[U_a = a_i]$ occurs' can operate as a sort of stand-in for the claim that $U_a = a_i$. In fact, the only difference between these two claims is that the former is not committed to the further claim that $\mathrm{H}[U_a]$ denotes a genuine cognitive variable.[7]

By the same process define $\mathrm{H}[U_b = b_i]$ and $\mathrm{H}[P = p_i]$. Given these definitions, proposition β_i clearly entails:

(Proposition γ_i) There is some hardware state of Eve's H_i, such that:

(i) State H_i is a member of $\mathrm{H}[U_a = a_i]$ and $\mathrm{H}[U_b = b_i]$ and $\mathrm{H}[P = p_i]$;
(ii) If Eve were to face conjunction E_i of external factors, then she would be in hardware state H_i;
(iii) If Eve were in hardware state H_i, then (regardless of the external factors E) she would choose C_i.

Which clearly entails:

(Proposition δ_i) There is some hardware state of Eve's H_i, such that:

(i) If Eve were to face conjunction E_i of external factors, then she would be in hardware state H_i;
(ii) if Eve were in hardware state H_i, then (regardless of the external factors E) she would choose C_i.

And this clearly entails proposition ϵ_i: if Eve were to face conjunction E_i of external factors, then she would choose C_i.

More importantly, the reverse entailments hold too. Proposition ϵ_i entails proposition δ_i, because external factors cannot cause choices except via hardware states. But proposition δ_i entails proposition γ_i, one can show.[8] And proposition γ_i entails proposition β_i, given one substantial assumption: $\mathrm{H}[U_a]$ constitutes a genuine cognitive variable, as does $\mathrm{H}[U_b]$ and $\mathrm{H}[P]$. And proposition β_i clearly entails proposition α_i. It follows that – given this substantial assumption about these three hardware/cognitive variables being genuine variables – proposition α_i is equivalent to proposition ϵ_i. But functionalist hypotheses are collections of propositions of type α. And collections of propositions of type ϵ are just hypotheses about input–output dispositions.[9] So any given two-step functionalist hypothesis is equivalent to (a) an hypothesis about an agent's input–output dispositions, plus (b) an hypothesis that says of some particular hardware/cognitive variables that they are genuine.

However, in effect, this logic shows that a two-step proposition (proposition α_i) is equivalent to a one-step proposition (proposition ϵ_i). And so by repeated application of this logic, a twenty-step proposition, say, might be shown to be equivalent to a one-step proposition – at least in some cases.[10] So the conclusion I've just drawn applies more generally to other functionalist hypotheses that contain multiple steps. So these functionalist hypotheses have psychological content – over and above the psychological content of an input–output disposition – in the respect that functionalist hypotheses make particular claims about which sets of hardware states constitute the values of genuine cognitive variable. So, if it turns out that it is very easy for a hardware/cognitive variable to be a genuine variable, then such functionalist hypotheses have little additional psychological content. But, if instead there are heavy-duty constraints on what counts as a genuine variable, then such functionalist hypotheses have lots of additional psychological content. This raises the third question in my framework: what is required in order for a set of hardware states to constitute the values of a genuine cognitive variable?

One necessary condition that I personally find attractive in this setting is the following: if a variable genuinely exists, U_a for example, then it is possible for that variable to take on different values from its actual value. For example, it is not possible, in any substantial sense, to intervene on Eve to make Eve a Porsche or a hanging basket or a horse. And, in virtue of this one might think, there is no 'is a Porsche' variable or a 'is a hanging basket variable' or a 'is a horse' variable that applies to Eve. Here I am talking about interventions in principle not interventions in practice, mind. This condition does not entail that there is a practically feasible intervention on Eve that would set the U_a variable to a different value. (To claim otherwise would be to place too demanding a constraint on the existence of a variable.)

A second plausible necessary condition is that the hardware states associated with a genuine variable all must be hardware states local to the same area of the brain; thus the only genuine cognitive variables are local. If this necessary condition is correct, it commits functionalist hypotheses such as H_F H_{WC} and H_{SC} to the following risky conjecture: utilities and probabilities are encoded in local regions of the brain; they are not features of the brain as a whole. The alternative is that utilities and probabilities are encoded by a huge complex of neural pathways that go all over the frontal lobe, the parietal lobe, the occipital lobe, the temporal lobe, and the limbic system, for example.

I think this condition is a good condition to impose if you are keen to ensure that functionalist hypotheses say something substantially different from hypotheses about input–output dispositions. But, of course, it's a bad condition to impose if you really don't care about what's going on at the hardware level, and so you don't want to commit to any hypotheses about what is going on at the hardware level, especially not any risky hypotheses. And not caring about what's going on at the hardware level is one of the key motivations for functionalism in the first place. So functionalists face a bit of a dilemma here, I think.

A third plausible necessary condition is that a variable is genuine only if the putative variable enters into some systematic counterfactual dependency relationships. But note that this necessary condition is automatically fulfilled, if proposition γ_i for example is fulfilled. So this is not much of a constraint at all in this setting.

A fourth plausible necessary condition says that idea of a variable being a genuine variable is a primitive notion that resists an easy definition. Such conditions are respectable within philosophy, but I suspect that most economists will not want to appeal to anything so metaphysical or mysterious.

In sum, I've argued that it's not clear what cognitive variables count as genuine variables. This motivates the third question in my framework: (III) does one intend to impose heavy-duty restrictions on what counts as a genuine hardware/cognitive variable? How one answers this question determines – surprisingly – how much of a gulf there is between behavioural scientists who aim to test functionalist hypotheses versus behavioural scientists who aim to test hypotheses about input–output dispositions. And this, in part, determines how much of a gulf there is between functionalist varieties of mentalism (on the one hand) and the behavioural tradition (on the other). This presents

functionalists with a dilemma: insofar as one places heavy-duty restrictions on what counts as a genuine cognitive variable, functionalist hypotheses make some risky claims about what's going on at the hardware level, something that functionalists want to avoid doing; but insofar as one doesn't place any heavy duty restrictions on what counts as a genuine cognitive variable, the functionalist position collapses to input–output dispositionalism, something more agreeable to the behavioural tradition from which functionalists want to distance themselves.

6. Does economics care about plasticity of choice dispositions?

A third way of interpreting the question 'should economics study the psychological basis of agents' choice behaviour?', I will now suggest, is to interpret it as the question 'should economics study whether agents' choice dispositions are plastic?'. It follows, as a corollary, that a helpful precursor to answering this question is to first ask question *IV* from my proposed framework: does economics ultimately aim to know whether agents' choice dispositions are plastic? In answering this question, I will raise a second problem for functionalist varieties of mentalism.

To understand this question, consider again the strongly conservative functionalist hypothesis H_{SC} from Section 4. That is to say, a modification of the hypothesis H_F by dropping all causal role requirements other than the means–end agency role. As I noted, H_{SC} does not predict what Eve's menus-to-choice mapping would have been, for example, if instead Eve had been socialised with people who listen to Beyonce (answer: Mapping Two).

Nevertheless, H_{SC} entails that it is possible for Eve's menus-to-choice mapping to have been different. In particular, there is some unspecified way of intervening on Eve that would have resulted in menus-to-choices Mapping Two, namely an unspecified intervention on variables U_a U_b that sets $U_b > U_a$.[11] This follows from the in-principle intervenability condition from Section 5 on a variable being a genuine variable. But, unlike functionalist hypothesis H_F, functionalist hypothesis H_{SC} doesn't entail anything specific about this intervention. It doesn't say of any specific socialisation and informational scenario, that in that scenario $U_b > U_a$ would be the case. But H_{SC} does entail that there exists some unspecified in-principle intervention on Eve that would bring out $U_b > U_a$, and would thereby bring about Mapping Two. (The same point goes for Mapping Three: H_{SC} entails that there exists some unspecified in-principle intervention on Eve that would bring out $P > .5$. and would thereby bring about Mapping Three. In contrast, H_F specifies what that intervention is: tell Eve that seventy percent of the cards in the deck are diamonds, for example.)

When one describes the menus-to-choices mappings that Eve could in principle have had (Mapping Two and Mapping Three for instance) – without saying anything about what specific interventions would bring about what mappings – I will say that one has given a non-specific description of the plasticity of Eve's menus-to-choice mapping. In this respect, even the strongly conservative H_{SC} describes non-specifically the plasticity of Eve's menus-to-choice mapping. (This distinguishes strong conservatism from extreme conservatism, which says that economics does not ultimately care about the relationship between any cognitive variables, not even the means-ends agency causal role. Economics does not ultimately care about any functionalist hypotheses.)

This peculiar feature of functionalist hypotheses has (to my knowledge) not before been pointed out. To see just how peculiar this feature is, consider the model of the consumer, interpreted as any form of functionalist hypothesis. According to this functionalist hypothesis, the consumer's choices over commodity bundles is plastic, I've just shown. The hypothesis says that the consumer's choice dispositions can be made to conform to any rank ordering one likes. There is a way of intervening on the consumer, for example, so that her choices conform to the weird rank ordering: 100 apples > 2 apples > 3 apples > 0 apples > 1 apple > 99 apples. Or, applying the model of the consumer to a voting context, there is a way of intervening on a voter so that her choices conform to the weird rank ordering: Centrist > Neo-Nazi > Free-market liberal > Stalinist > Green > Social democrat. I myself find it somewhat implausible that any consumer's or voter's choices are that plastic

(malleable). So any functionalist interpretation of the model of the consumer interprets this model as making some very claims, I suggest, about the plasticity of consumer's choice dispositions.

Corollary. This illustrates the fourth question in my proposed framework: does economics ultimately care about knowing non-specifically the plasticity of Eve's choices? For example, imagine that you believe a functionalist hypothesis about how voters are socialised into having utilities over political ideologies, and how they use information to form beliefs about which candidates represent which political ideologies. But suppose that you then discover the following. Yes, your hypothesis is exactly right about who a voter would vote for if she were to receive different information, or if she had been socialised differently. In this respect, your model correctly describes the plasticity of her vote with respect to these two specific external factors. But, no, your hypothesis is wrong in another respect: since your hypothesis is a functionalist hypothesis, it has the peculiar and risky implication that there is, in principle, some unspecified intervention on this voter's hardware states that would set u(Centrist) > u(Neo-Nazi) > u(Free-market liberal) > u(Stalinist) > u(Green) > u(Social democrat); and this implication is false, you discover. In this respect, your model incorrectly describes the non-specific plasticity of her vote. Does economics care about this discovery as an end in itself? Does this discovery about this voter speak to one of the questions about the voter that economics was ultimately aiming to answer? It is clear that Gul and Pesendorfer (2008) would answer no, as would Clarke (2016) and Thoma (2020). Dietrich and List (2016) say that economics ultimately aims to test functionalist hypotheses, and so they are implicitly committed to defending a yes answer. What about other mentalists, such as Hausman (2012)? On the one hand, most philosophers agree that functionalism is the most plausible philosophy of mind, and so Hausman and others have strong reason to answer yes. But, on the other hand, functionalism when applied to economic models has the risky implication that I've just pointed out. So Hausman and others also have strong reason to answer no. Another dilemma.

Why is it a good thing to answer question *IV*? It's a good thing, because how you answer this fourth question determines what evidence is relevant for economics. Take for example Glimcher's experiments that purport to show that, for each option in a choice experiment, there are a highly-localised bunch of neurons that is causally responsible for the agent's inclination to choose that option over other options (Glimcher et al., 2005). Glimcher's neural observations are some evidence that it is possible to intervene on economic agents such that they exhibit a very different pattern of choice behaviour: just find some way of intervening on these neurons. So Glimcher's neural observations provide some evidence that agents' choice dispositions are very plastic. Therefore, Glimcher's neural observations provide some evidence for any functionalist hypothesis, since all functionalist hypotheses are committed to Eve's choice dispositions being very plastic. So, if the answer to question *IV* is yes, then Glimcher's observations are relevant to economics.

I conclude that a third way of interpreting the question 'should economics study the psychological basis of agents' choice behaviour?' is to interpret it as the question 'should economics study non-specifically the plasticity of agents' menu-to-choice mapping?'. It followed, as a corollary, that a helpful precursor to answering this question is to first ask question *IV* from my proposed framework: does economics ultimately care about knowing non-specifically the plasticity of these mappings?

7. Does economics care about intentional choice?

In this section, I will point out that the notion of choice behaviour is also ambiguous. On the one hand there is intentional choice, and on the other hand there is what I will call eliminativist choice. It follows, as a corollary, that a helpful precursor to answering the psychology-in-economics question – about whether economics should study the psychological basis of choice behaviour – is to first ask question *V* from my proposed framework: does economics ultimately aim to predict and explain an agent's choices qua intentional choices? Since this point has already been made briefly by Clarke (2016) and more carefully and thoroughly by Thoma (2020), my treatment here will be quick.

Recall the model from Section 2, which refers, amongst other things, to the event of Eve choosing to hear Aerosmith guaranteed. Note that there are two ways of theorising events such as this one. One theory says that this event is the event of Eve's body moving in such a way that she actually did hear Aerosmith. An alternative theory says that this event is the event of Eve *intentionally acting* in such a way that Eve *knew* would cause her to hear Aerosmith. To see the difference between these two theories of choice events, imagine that Eve walks into a cafe and inadvertently ends up hearing Beyonce; but, if she had walked into a second cafe, she would have instead inadvertently heard Aerosmith. On the former theory, Eve has chosen to hear Beyonce over Aerosmith. But on the latter theory, she has not. I will call the former theory the 'eliminativist' theory of choice events, because it eliminates the vocabulary of commonsense psychology from our theory of choice events. Choice events are not theorised in terms of an agent's intentional actions or an agent's knowledge. Instead, the theory says that for an agent A to choose an option O_1 over an option O_2 is just for A's body to move in such a way as O_1 occurs rather than O_2.

In contrast, I will call the alternative theory the 'ordinary intentional' theory of choice events, because it's the understanding we typically adopt in everyday and philosophical discussions. Choice events are theorised in terms of an agent's intentional actions or knowledge or similar. To be specific, the ordinary intentional theory says that for an agent A to choose an option O_1 over an option O_2 is just for the following to hold: (i) A forms an intention to move her body in way X; and (ii) A knows that moving her body in way X will result in O_1 occurring rather than O_2 occurring.[12] For more on the distinction between intentional action and mere bodily movement see Wilson and Shpall (2016); see also Thoma (2020) for a discussion in the context of economic models of choice behaviour.

Next, note that Mapping One is at least part of the content of the model from Section 1, everyone agrees:[13]

> If Eve's menu were any possible subset of the four options *I II A B*, then Eve would choose the highest option on the ordering *A* > *I*>*II*>*B*.

It follows that, on the eliminativist theory of choice behaviour, the model has the following as part of its content: Eve's body never moves in a way that results in her hearing Beyonce, when Eve's body could have moved in a way that results in her hearing Aerosmith. As Hausman (2012) points out, this claim is very rarely true of an agent: sometimes Eve is ignorant of the effects of her actions and so Eve's body does indeed move in a way that results in her hearing Beyonce. Eve inadvertently wanders into a cafe in which Beyonce is playing, for example. So on the eliminativist theory of choice behaviour the content of most economic models of choice behaviour is false. I call this Hausman's Objection from Ignorance.

Note, however, that this problem does not arise if one theorises choice behaviour in the ordinary intentional way. On the ordinary intentional theory of choice behaviour, the model has the following as part of its content: Eve never intentionally acts in such a way that she knows will result in her hearing Beyonce, when she could intentionally act in a way that she knows will result in her hearing Aerosmith. This is a plausible claim to make of Eve. So Hausman's Objection from Ignorance does not arise for models on a ordinary intentional interpretation of choice, even though it does for models on an eliminativist interpretation of choice.

(This distinction between the ordinary intentional theory and eliminativist theory generalises: one might make exactly the same point about the external factors to which the hypothesis H_F refers: Eve being socialised with people who like Aerosmith, Nina telling Eve that n out of the m cards are diamonds, and the like. One can either read these conditions in the ordinary sense – as making claims about Nina's intentions for example. Or one can try to give them an eliminativist reading.)

I conclude that the notion of choice behaviour is ambiguous between intentional choice and eliminativist choice. It follows, as a corollary, that a helpful precursor to answering the question – about whether economic should study the psychological basis of choice behaviour – is to first ask question

V from my proposed framework: does economics ultimately aim to predict and explain an agent's choices qua intentional choices?

8. Conclusion

I've argued that the controversy over the proper role of psychological concepts, explanations and evidence in economics is multifaceted and profoundly indeterminate. Firstly, the notion of choice behaviour is ambiguous, in a way that blurs the boundaries between the mentalist and behavioural traditions (Section 7). Secondly, the notion of a cognitive variable is ambiguous, in a way that threatens to collapse psychological hypotheses about cognitive processes into hypotheses about input–output dispositions, hypotheses more agreeable to the behavioural tradition (Section 5). This supports Guala's contention that 'terms like mental and psychological, unfortunately, are often used differently by economists, psychologists and philosophers, generating considerable confusion in this debate' (Guala, 2019, p. 387).

Thirdly, there are at least three parts to the mentalist behavioural controversy. First, there is the issue of whether economics should study the brain anatomy or hardware that realises agents' functional–dispositional states (Section 3). A second part of this controversy is the question of whether economics should study the causes and effects of cognitive variables (Section 4). A third part of this controversy, I suggest, is whether economics should study non-specifically the plasticity of agents' choice dispositions (Section 6).

One benefit of clearing up these ambiguities in this controversy is that it throws light on some problems for functionalist varieties of mentalism (Dietrich & List, 2016). Section 5 showed that functionalists face a dilemma: insofar as one places heavy-duty restrictions on what counts as a genuine cognitive variable, functionalist hypotheses make some risky claims about what's going on at the hardware level, something that functionalists don't want to do; but insofar as one doesn't place any heavy duty restrictions on what counts as a genuine cognitive variable, the functionalist position collapses to a form of dispositionalism, something more agreeable to the behavioural tradition from which functionalists distance themselves. What's more, Section 6 showed that, on even a very modest restriction on what counts as a genuine cognitive variable, functionalist hypotheses make some implausible commitments to the plasticity of agent's options-to-choices mappings.

En route, I've argued that part of the controversy is about the ultimate aims of economics as a discipline, and that it's unhelpful to vaguely assert that economics ultimately aims to predict and to explain choice behaviour (Section 2). Instead, I've suggested a framework of five useful questions:

(I) Does economics ultimately care about the brain anatomy or hardware that realise any functional–dispositional states?
(II) Which external causes, which cognitive causes, and which cognitive effects of cognitive variables does economics ultimately care about identifying, if any?
(III) Does one intend to impose heavy-duty restrictions on what counts as a genuine cognitive variable?
(IV) Does economics ultimately care about knowing non-specifically the plasticity of agents' menus-to-choices mapping?
(V) Does economics ultimately aim to predict and explain an agent's intentional choices, as opposed to her choices in an eliminativist sense?

The five questions in my framework allow one to make it clear what one takes economics to be ultimately aiming at. As such, my framework makes it clear how one can develop various different positions that occupy a middle ground in between the mentalist and behavioural traditions. For example, Clarke (2014, 2016), Thoma (2020) and Vredenburgh (2020), give a mentalist answer to

question *V*, but a behavioural answer to questions *I* and *II*. Therefore, I hope that my framework will allow for a much clearer discussion of the controversy: what sort of neurobiological (brain scanner), cognitive psychological (reaction time), or phenomenological (self-report) evidence can be used to test and improve the sorts of hypotheses that economics ultimately aims to test?

Notes

1. For philosophical defences of this answer see Wong (1978/2006), Rosenberg (1988/2008), Sen (1993), Rosenberg (1992), Cohen (1995), Camerer et al. (2005), Craver and Alexandrova (2008), Alexandrova and Haybron (2011), Lehtinen (2011), Hausman (2012), Reiss (2013), and Dietrich and List (2016).
2. See Samuelson (1948, p. 251) and Gul and Pesendorfer (2008) as extreme statements of the ethos of the behavioural tradition. For less extreme statements, see Friedman and Savage (1952), Binmore (2007a, p. 321), Binmore (2009, p. 14), Bernheim and Rangel (2008), Wakker (2010, p. 366), and Gilboa (2010, p. 20), as well as many textbooks, including Mas-colell et al. (1995, p. 11), Varian (2005, p. 120), Varian (1992, pp. 131–133), Binmore (2007b, Chapters 1.4.2, 14.1–14.2).
3. Although see Section 7 for a respect in which even Mapping One might have some psychological content. And see Clarke (2014) for extensive discussion.
4. For agreement with this emphasis on external factors as causes of choice, see Friedman and Savage (1948), Bernheim (2008, p. 3), Binmore (2009), Dekel and Lipman (2010, p. 273), and Vromen (2011, sec. 6).
5. See any of the discussions within meta-ethics on moral motivation (Smith, 1994) or within the philosophy of action on reasons for action (Velleman, 2000). The discussion on 'eliminative materialism' in the philosophy of mind can also be read as a debate about to what extent any neural variables actually do play the causal roles that define commonsense psychology (Churchland, 1981; Horgan & Woodward, 1985; Jackson & Pettit, 1990).
6. More accurately, they are this long list prefaced with 'There exist variables U_a and U_b and P such that . . .'.
7. That is to say, the further claim that {A state in H[$U_a = a_1$] occurs, A state in H[$U_a = a_2$] occurs, A state in H [$U_a = a_3$] occurs, . . . } constitutes the values of a genuine cognitive variable.
8. Proposition δ_i says that E_i caused H_i. But the functionalist hypothesis says that E_i causes $U_a = a_i$ and $U_b = b_i$ and $P = p_i$. And so, our definition of H[$U_a = a_i$] and of H[$U_b = b_i$] and of H[$P = p_i$] tells us that H_i is a member of all three sets.
9. To be contrasted, for example, with Guala's talk of 'belief-dependent' dispositions.
10. This is just a brief outline of an argument. I suspect there will be a number of important qualifications to the scope of this argument. Firstly, note that the functionalist hypotheses I am dealing with say nothing about the relative timing of changes in values of the cognitive hardware variables. Secondly, I suspect that cognitive/hardware variables that make a repeat appearance in a functionalist hypothesis (e.g. at step two and then again at step ten) will pose difficulties for this argument.
11. This does specify $U_b > U_a$ but says nothing about what that would involve in hardware terms.
12. In fact, this is a highly simplified version of the theory, about which there is much philosophical controversy (Wilson & Shpall, 2016).
13. Some economists and philosophers would say that Mapping One exhausts the content of the model, whereas others think that there is more to the content of the model than Mapping One alone.

Acknowledgements

Thank you Chloe de Canson, Caglar Dede, Conrad Heilmann, Francesco Guala and two anonymous referees for your gracious and constructive comments on the manuscript. I'm also indebted to Johanna Thoma and Kate Vredenburgh, and to the participants at their 2018 workshop at the LSE on Revealed Preferences, for discussion of some of these ideas.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work has received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation Programme, under grant agreement no 715530.

Notes on contributor

Christopher Clarke is a Senior Research Associate at the University of Cambridge (CRASSH) and an Assistant Professor at Erasmus University Rotterdam (School of Philosophy). He works on the nature of causal explanation and causal inference, especially in political science and economics.

ORCID

Christopher Clarke http://orcid.org/0000-0002-6225-0115

References

Alexandrova, A., & Haybron, D. M. (2011). High fidelity economics. In J. Davis & W. Hands (Eds.), *Elgar companion to recent economic methodology* (pp. 94–120). Edward Elgar. https://doi.org/10.4337/9780857938077.00010

Bernheim, B. D. (2008). *Neuro-economics: A sober (but hopeful) appraisal* (NBER Working Paper Series, pp. 1–53). http://www.nber.org/papers/w13954

Bernheim, B. D., & Rangel, A. (2008). Choice-theoretic foundations for behavioral welfare economics. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 155–192). Oxford University Press.

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

Binmore, K. (2007b). *Playing for real: A text on game theory*. Oxford University Press.

Binmore, K. (2009). *Rational decisions*. Princeton University Press.

Block, N. (1980). Introduction: What is functionalism? In N. Block (Ed.), *Readings in the philosophy of psychology* (Vol. 1, pp. 171–184). Harvard University Press.

Camerer, C. F. (2008). The case for mindful economics. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 43–69). Oxford University Press.

Camerer, C. F., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*(1), 9–64. https://doi.org/10.1257/0022051053737843

Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford University Press.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, *78*, 67–90. http://www.jstor.org/stable/2025900

Clarke, C. (2014). Neuroeconomics and confirmation theory. *Philosophy of Science*, *81*(2), 195–215. https://doi.org/10.1086/675669

Clarke, C. (2016). Preferences and positivist methodology in economics. *Philosophy of Science*, *83*(2), 192–212. https://doi.org/10.1086/684958

Cohen, J. (1995). Samuelson's operationalist–descriptivist thesis. *Journal of Economic Methodology*, *2*(1), 53–78. https://doi.org/10.1080/13501789500000003

Craver, C., & Alexandrova, A. (2008). No revolution necessary: Neural mechanisms for economics. *Economics and Philosophy*, *24*(3), 381–406. https://doi.org/10.1017/S0266267108002034

Davidson, D. (1984). Radical interpretation. In *Inquiries into truth and interpretation* (pp. 125–140). Clarendon–OUP. (Original work published 1973).

Dekel, E., & Lipman, B. L. (2010). How (not) to do decision theory. *Annual Review of Economics*, *2*(1), 257–282. https://doi.org/10.1146/annurev.economics.102308.124328

Dietrich, F., & List, C. (2016). Mentalism versus behaviourism in economics: A philosophy-of-science perspective. *Economics and Philosophy*, *32*(2), 249–281. https://doi.org/10.1017/S0266267115000462

Elster, J. (1985). *Sour grapes: Studies in the subversion of rationality* (Reprint ed.). Cambridge University Press.

Eriksson, L., & Hájek, A. (2007). What are degrees of belief? *Studia Logica*, *86*(2), 183–213. https://doi.org/10.1007/s11225-007-9059-4

Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Bradford–mit.

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy*, *56*(4), 279–304. https://doi.org/10.1086/256692

Friedman, M., & Savage, L. J. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy*, *60*(6), 463–474. https://doi.org/10.1086/257308

Gilboa, I. (2010). *Rational choice*. MIT Press.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior*, *52*(2), 213–256. https://doi.org/10.1016/j.geb.2004.06.011

Guala, F. (2019). Preferences: Neither behavioural nor mental. *Economics and Philosophy*, *35*(3), 383–401. https://doi.org/10.1017/s0266267118000512

Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics: A handbook* (pp. 3–39). Oxford University Press.

Hausman, D. M. (2000). Revealed preference, belief and game theory. *Economics and Philosophy*, *16*(1), 99–115. https://doi.org/10.1017/S0266267100000158
Hausman, D. M. (2012). *Preference, value, choice and welfare*. Cambridge University Press.
Horgan, T., & Woodward, J. (1985). Folk psychology is here to stay. *The Philosophical Review*, *94*(2), 197–226. https://doi.org/10.2307/2185428
Jackson, F., & Pettit, P. (1990). In defence of folk psychology. *Philosophical Studies*, *59*(1), 31–54. https://doi.org/10.1007/BF00368390
Lehtinen, A. (2011). The revealed-preference interpretation of payoffs in game theory. *Homo Oeconomicus*, *28*, 265–296.
Lewis, D. K. (1983). Mad pain and martian pain. In *Philosophical papers* (Vol. 1, pp. 122–129). Oxford University Press. (Original work published 1980).
Lewis, D. K. (1999). Reduction of mind. In *Papers in metaphysics and epistemology* (pp. 291–324). Cambridge University Press. (Original work published 1994).
List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.
Mas-colell, A., Winston, M. D., & Green, J. R. (1995). *Microeconomic theory*. Oxford University Press.
Meacham, C. J. G., & Weisberg, J. (2011). Representation theorems and the foundations of decision theory. *Australasian Journal of Philosophy*, *89*(4), 641–663. https://doi.org/10.1080/00048402.2010.510529
Reiss, J. (2013). *Philosophy of economics: A contemporary introduction*. Routledge–Taylor.
Rosenberg, A. (1992). *Economics: Mathematical politics or science of diminishing returns?* University of Chicago Press.
Rosenberg, A. (2008). *Philosophy of social science* (3rd ed.). Westview. (Original work published 1988).
Ross, D. (2005). *Economic theory and cognitive science*. Bradford–mit.
Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, *15*(60), 243–253. https://doi.org/10.2307/2549561
Sen, A. K. (1993). Internal consistency of choice. *Econometrica*, *61*(3), 495–521. https://doi.org/10.2307/2951715
Shoemaker, S. (1984). Some varieties of functionalism. In *Identity, cause, and mind: Philosophical essays* (pp. 261–286). Oxford University Press. (Original work published 1981).
Smith, M. (1994). *The moral problem*. Blackwell.
Thoma, J. (2020). In defence of revealed preference theory. *Economics and Philosophy*, 1–25. https://doi.org/10.1017/s0266267120000073
Varian, H. R. (1992). *Microeconomic analysis* (3rd ed.). Norton.
Varian, H. R. (2005). *Intermediate microeconomics: A modern approach* (7th ed.). Norton.
Velleman, D. (2000). *The possibility of practical reason*. Oxford University Press.
Vredenburgh, K. (2020). A unificationist defence of revealed preferences. *Economics and Philosophy*, *36*(1), 149–169. https://doi.org/10.1017/s0266267118000524
Vromen, J. (2011). Neuroeconomics: Two camps gradually converging: What can economics gain from it? *International Review of Economics*, *58*(3), 267–285. https://doi.org/10.1007/s12232-011-0127-8
Wakker, P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge University Press.
Wilson, G., & Shpall, S. (2016). Action. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016). http://plato.stanford.edu/archives/win2016/entries/action/; Metaphysics Research Lab, Stanford University.
Wong, S. (2006). *The foundations of Paul Samuelson's revealed preference theory* (2nd ed.). Routledge. https://doi.org/10.4324/9780203462430. (Original work published 1978).
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
Zynda, L. (2000). Representation theorems and realism about degrees of belief. *Philosophy of Science*, *67*(1), 45–69. https://doi.org/10.1086/392761