CrossMark

Economics and Philosophy (2019), **35**, 383–401
doi:10.1017/S0266267118000512

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Preferences: neither behavioural nor mental

Francesco Guala

Department of Philosophy, Università degli Studi di Milano, via Festa del Perdono 7, 20136 Milano, Italy
Email: francesco.guala@unimi.it.

(Received 4 April 2017; revised 24 September 2018; accepted 28 September 2018; first published online 6 March 2019)

Abstract
Recent debates on the nature of preferences in economics have typically assumed that they are to be interpreted either as behavioural regularities or as mental states. In this paper I challenge this dichotomy and argue that neither interpretation is consistent with scientific practice in choice theory and behavioural economics. Preferences are belief-dependent dispositions with a multiply realizable causal basis, which explains why economists are reluctant to make a commitment about their interpretation.

Keywords: preference; choice; behavioural economics; dispositions; mental states

1. Introduction

Some philosophical controversies never die. The debate about the interpretation of preferences and utility[1] in economics flared up at the end of the 19th century;[2] in the 1940s and 1950s economists seemed to converge on a shared interpretation, but the consensus was fragile: with the rise of behavioural and experimental economics, the controversy has restarted forcefully during the past decade.[3] The main bone of contention is whether preferences and utility should be given a *behavioural* or *mental* interpretation. Do these theoretical constructs refer to overt behaviour, patterns of choice that are readily observable and intersubjectively measurable? Or should we take them to describe psychological states, hidden in the minds of decision-makers?

These are generally assumed to be the only plausible interpretations. But the assumption is dubious: in the next few pages I will make a proposal that cuts across the behavioural/mental dichotomy and tries to reconcile the interpretation of

[1]Although I will use the term 'preference' more frequently, I will take preferences and utility to be roughly synonymous, in line with the standard terminology of contemporary choice theorists (a utility function is an index of preferences).

[2]Historical reconstructions of key aspects of the old controversy can be found in Lewin (1996), Bruni and Sugden (2007), Edwards (2016), Moscati (2018, Forthcoming).

[3]An incomplete list of contributions to the recent debate includes Hausman (2000, 2008, 2012), Mongin (2000a), Dowding (2002), Binmore (2008), Glimcher *et al.* (2005), Ross (2005, 2011, 2014), Camerer (2008), Gul and Pesendorfer (2008), Hands (2009, 2013, 2014), Vromen (2010), Lehtinen (2011), Guala (2012), Fumagalli (2013), Cozic and Hill (2015), Clarke (2016), Dietrich and List (2016), Okasha (2016), Engelen (2017), Angner (2018), Nagatsu and Poder (Forthcoming).

© Cambridge University Press 2019

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

preferences with the best scientific practice in economics. I will argue that the two standard positions in this debate – 'behaviourism' and 'mentalism' – are to be rejected. Preferences in the economic theory of choice are dispositions that can be realized in different ways depending on the circumstances of choice and on the characteristics of the decision-maker. I will argue that this claim does not imply a commitment to a behaviourist interpretation of preferences – on the contrary it is compatible with the idea that psychology plays an important role in choice theory. But it does not imply that preferences are to be transformed into psychological constructs, in spite of what some economists advocate (and others dread).

The argument relies crucially on multiple realization, and on the 'wide' applicability of choice theory beyond the domain of human individual decision-making. This in turn requires that 'preference' is given an abstract interpretation, along the same lines as concepts like 'force' in physics or 'fitness' in biology. In physics the discovery that there are different kinds of forces prompted the development of different theories, such as electromagnetism, that now complement the traditional theory of gravitation. Similarly, different theories of preferences ought to explain the preferences of decision units whose internal processes are unlikely to be explicated by the principles of human psychology.

The discussion will proceed as follows: Section 2 summarizes classic behaviourism and explains why it is untenable. This section is relatively short and unoriginal. It will outline what I take to be the fundamental flaw of behaviourism, refer to the relevant literature, and save space for the discussion of mentalism in the subsequent sections. Section 3 introduces a prominent approach to the identification of mental states – functionalism – and explains why it fails to capture economists' concerns about the psychologization of choice theory. Section 4 explains what role preferences play in economic explanations of behaviour, in dispositional terms. Section 5 shows that this interpretation is consistent with scientific practice in behavioural economics, focusing on the case of Prospect Theory. The key argument against the behavioural-mental dichotomy is presented in Section 6, where I illustrate how intransitive preferences can be multiply realized, using examples from human decision-making and committee deliberation. Section 7 summarizes the argument and concludes the paper.

2. Background: why preferences are not behaviour

The interpretation of preferences and utility that became orthodox in the middle of the 20th century is a direct descendant of philosophical 'positivism'. Positivism, to be sure, took various forms: while bland positivists merely stressed the importance of empirical evidence for the production of knowledge, radical ones advocated the elimination of all scientific terms that are directly unobservable or irreducible to observable concepts. In physics, for example, operationalists tried to reduce the meaning of theoretical statements to the set of actions that scientists perform when they measure a parameter in the laboratory (e.g. Bridgman 1927). An electron according to this interpretation is not a small particle with a negative electric charge, but the set of operations we perform when we observe, say, tracks in a bubble chamber. In psychology, behaviourists aimed at re-interpreting every psychological

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

concept in terms of measurable behaviour (e.g. Watson and McDougall 1929). Terms like 'hunger' or 'pain' were supposed to be replaced by laws that connect external physical stimuli (like food deprivation) with overt behaviour (like food seeking).

Similar ideas played a prominent role in choice theory, where positivistic-minded economists attempted (1) to replace 'psychophysical' utility (pleasure) with an index of preferences; and (2) to reduce preferences to observed choice (behaviour). The first part of the programme was accomplished by various theorists (notably Pareto 1906; Hicks 1939) over the course of thirty years, and has become a pillar of modern microeconomics. The second part, in contrast, proved to be more problematic right from the beginning, and led to ambiguities that still hamper contemporary discussions on the status of preferences.

The best known application of behaviourism in economics is Paul Samuelson's theory of 'revealed preferences' (Samuelson 1938, 1948, 1950). Samuelson's work was partly motivated by his dissatisfaction with Pareto's and Hicks' preference theory. Samuelson presented his proposal as a 'direct attack upon the problem, dropping off the last vestiges of the utility analysis' (Samuelson 1938: 62). Taking choice as primitive, he showed that the standard analysis of consumer's behaviour (Neoclassical demand theory) only requires a simple restriction (the 'weak' axiom of revealed preferences) on agents' choices: 'if an individual selects batch one over batch two, he does not at the same time select two over one' (Samuelson 1938: 65). Samuelson thus aimed at obtaining, without mentioning utility or preferences, the same results that had been achieved by Pareto.[4]

Revealed preference theory was refined and developed in various directions during the following two decades,[5] and Samuelson's framework has become a standard tool in advanced economic textbooks (e.g. Mas-Colell *et al.* 1995; Varian 2005). But in spite of its apparent success, the original behaviourist project was never fully achieved. There are many reasons for this,[6] but the most fundamental one has to do with its narrow original scope: Samuelson's revealed preference approach was meant to be a contribution to riskless decision-making. The formal mapping of preferences onto behaviour was attained in a belief-less context, in other words, taking for granted that consumers have perfect information about the objects of choice. But if the latter condition is dropped, it is easy to demonstrate that there cannot be a one-to-one correspondence between preference and choice.

Consider the following two cases:

(a) Tony prefers the restaurant Pizza Vesuvio to Pizza Bella Napoli. He believes that Pizza Vesuvio is closed tonight. Therefore, he goes to Pizza Bella Napoli.

[4]This behaviourist interpretation of Samuelson's project, which was once majoritarian, has been the target of sustained criticism in recent years. Mongin (2000a) for example argues that Samuelson never meant to replace ordinal utility analysis with the theory of revealed preferences, while according to Moscati (Forthcoming) he was never a committed behaviourist.

[5]See e.g. Samuelson (1948), Houthakker (1950), Afriat (1967). Hands (2013) proposes a useful taxonomy of different versions of the revealed preference project.

[6]Caplin (2008) explains why and how the convergence of ordinal utility and revealed preference theory (a 'paradise lost') has collapsed. For a comprehensive reconstruction of the evolution of Samuelson's position, see Hands (2014).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

(b) Vincent prefers Pizza Bella Napoli to Pizza Vesuvio. He believes that both are open. Therefore, he goes to Pizza Bella Napoli.

Although their choices are identical, it would be a mistake to conclude that Tony and Vince have the same preferences. The reason is that, in the standard theory of economic choice, behaviour is determined both by preferences and by beliefs. As a consequence, the same choice (behaviour) may result from different preferences, if beliefs also differ.

The problem is ubiquitous. Tony may decide not to contribute to the organization of the Christmas fair because he believes that no one in the neighbourhood will give any money, although he regrets that the fair will not take place. Vince in contrast does not contribute because he hates Christmas fairs and wouldn't give any money even if everyone else did. By looking at their behaviour we cannot determine whether the situation that each of them is facing is a public goods game, a coordination game, or something else, because we do not know how their choices depend on their beliefs about other people's preferences and beliefs.

One option, of course, is to add beliefs to the background conditions that allow the mapping of preferences onto choices. Such a move is eminently sensible and scientifically correct, but defies the behaviourist goal: unless the beliefs are defined in behavioural terms, the revealed preference theorist would make use of the sort of psychological concepts that she intended to eliminate in the first place. And exactly the same argument holds, symmetrically, for beliefs: it is possible to derive beliefs from behaviour only if we know the preferences of decision-makers. The behaviourist is trapped in a methodological circle that does not have a satisfactory way out.

This point has been thoroughly made by Alex Rosenberg (1993), Dan Hausman (2000, 2012), and others, so I will not devote much time to it here. Some economists have proposed interpretations of choice theory that attempt to circumvent the problem in ingenious ways, but none of them is able to recover the behaviourist programme without betraying its original goal.[7] Without informational input about 'primitive' variables (i.e. variables that are irreducible to behaviour), it is impossible to theorize in core areas of economics such as decision and game theory.

An alternative strategy is to abandon radical behaviourism and endorse a weaker interpretation of revealed preference theory. This weaker but more reasonable interpretation has always co-habited with the radical behaviourist one: instead of claiming that preferences *are* behaviour, many economists take Samuelson to mean that preferences are *manifested* in behaviour. The latter view is suggested by Samuelson's use of the term 'revealed',[8] and is consistent with a moderate empiricist attitude: although preferences and choices are different things, choice is the empirical evidence that economists traditionally use to test hypotheses derived from beliefs, desires and option sets.[9]

[7]See Binmore (1994), Hausman (2000, 2012) and Hands (2013) for discussion. Dietrich and List (2016) review and discuss several other arguments against the behaviouristic interpretation of preferences.

[8]The terminology was introduced in Samuelson (1948). The expression 'revealed preference' does not even occur in Samuelson's first (1938) article.

[9]This distinction between a strong and weak form of revealed preference theory is now entrenched in the literature: see e.g. Hausman (2000, 2012), Hands (2013), Dietrich and List (2016), and, for behaviourism more generally, also Moore (2001).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

But endorsing the moderate interpretation shifts the problem one step beyond. What is choice behaviour evidence *of*? If preferences are not patterns of choice, then what are they? What is the alternative to a behaviourist interpretation? Does rejecting behaviourism imply that we should accept a 'mentalistic' interpretation of preferences?

3. The riddle of the mental

Most economists do not care much about behaviourism, to be sure. Contemporary supporters of revealed preference theory are more interested in raising barriers against the intrusion of psychology in economics, than in reviving an out-dated philosophical programme.[10] Most of them however share a common presumption, namely, that the failure of behaviourism would open the doors to mentalism in economics. Terms like 'mental' and 'psychological', unfortunately, are often used differently by economists, psychologists and philosophers, generating considerable confusion in this debate. The first step therefore is to clarify what a mentalistic interpretation of preference could possibly be.

Most contemporary philosophers have abandoned the idea that mental states can be identified with intrinsic properties, such as their experiential content. In contrast, they believe that mental states should be identified with the role they play in the aetiology of behaviour. In a recent paper Franz Dietrich and Christian List describe this approach as follows:

> Mental states are, at least in part, states that play a certain role for an agent. Beliefs, for example, play the role of representing certain features of the world from the agent's perspective, and preferences play the role of motivating the agent's actions ... . *Functionalism* is the view that what makes something a mental state is simply that it plays the relevant role. (Dietrich and List 2016: 268)[11]

Functionalism (or 'non-reductive materialism', as it is sometimes called) offers a straightforward solution to the problem of preferences: they are mental states because they play the 'right' motivational role in human action. Assuming a deflated notion of the mind, functionalists like Dietrich and List tell economists that, unknowingly, they have been talking about mental states all along.

Such a solution however looks suspiciously simple: is the mentalist-behaviourist controversy *really* based on a trivial misunderstanding? Can we solve the disagreement

[10]Gul and Pesendorfer (2008) are typical in this respect. Another concern is to preserve the (alleged) value-neutrality of economic science; on the interpretation of preferences and the normative/positive uses of choice theory, see e.g. Hausman (2012), Hands (2013: sec. 4) and Okasha (2016). Sen (1982) is the precursor of most contemporary discussions on the use of preferences in normative economics.

[11]Dietrich and List endorse a weak form of functionalism, according to which playing a preference-role is indicative, rather than constitutive, of mental states. The main reason is that mental states may have other, non-functional characteristics that can only be ascertained from a first-personal perspective. A stronger functionalist position is defended by List and Pettit (2011), who take exclusively a third-personal perspective on the behaviour of groups. I'm grateful to Franz Dietrich and Christian List for clarifying this point.

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

merely by redefining the notion of mental state? If mentalism really meant so little, why the fuss? Why do many economists resist it, and why do others advocate it so fervently?

The controversy is only partly terminological. Of course one can *define* preferences as mental states in the functionalist sense, but the terminological issue has some important substantial ramifications. A functionalist definition, to begin with, fails to capture the reason why mentalism is controversial in economics: behavioural economists do not merely promote the reinterpretation of a key economic concept – they advocate a robust injection of *psychological theory and methods* in the economics of individual choice. Those who resist this programme, in contrast, believe that the theory of choice should be kept separate from the science of human psychology. In both cases the debate concerns substantial scientific issues rather than conceptual or definitional worries about the mental.

Functionalists like Dietrich and List take preference-based explanations to be distinct from lower-level explanations that appeal to the underlying realizers of those preferences. Although I share this fundamental premise, I will defend an alternative interpretation of preference theory, based on a different conception of what a science of the mental (or a science of psychology) should do. Like most economists, I will take psychology to be engaged in the *explanation of human preferences*, rather than in the explanation of behaviour by means of preferences (which is the specific task of choice theory). Moreover, I will argue that no psychological explanation of preferences is likely to cover all the domains in which choice theory is applied.

I share with Dietrich and List a broadly naturalistic attitude: any acceptable interpretation of choice theory must be consistent with scientific practice. Our disagreement therefore stems partly from a different understanding of the mentalist project in economics. I will argue that behavioural economists and psychologists of decision-making have not rejected economists' traditional conception of preferences as dispositional states. They have rather challenged the traditional properties of preference rankings, and have tried to convince us that human utility functions have certain peculiar shapes in virtue of certain psychological mechanisms that govern the formation of human preferences. Mentalism is first and foremost a scientific, explanatory project, rather than an interpretation of choice theory.

Classic functionalism, to be sure, was also rooted in a scientific programme. The main goal of this programme – 'Artificial Intelligence' – was to understand cognition by modelling the mind as a computational device. A cognitive system, according to this approach, is essentially a software that can be implemented in many different ways. The hardware does not matter – there can be minds made of silicon chips, vacuum tubes, or flesh and blood. As long as 'something' plays the right role in their control systems, non-human animals, robots, plants and aliens can be legitimately (scientifically) said to have mental states.[12]

Functionalism aimed at convincing philosophers and scientists that it is possible to understand complex cognition without 'looking under the hood'. More concretely, it promised to understand the mind without a full understanding of the way the brain works. The software can be studied independently of its hardware.

[12]For a classic statement, see for example the essays on the mind-body problem collected in Putnam (1975).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

Thus functionalism was motivated primarily by a *methodological*, rather than by an ontological concern. Methodological concerns are also at the root of the choice theory controversy. But paradoxically they are the flip image of what the functionalist approach would suggest. The main goal of behavioural economists has been to persuade social scientists that it *is* important to 'look under the hood' – that is, to understand the psychological basis of economic decision-making. If there is a methodological commitment behind mentalism in economics, it is exactly the opposite than the functionalist's commitment in AI and the philosophy of mind.

The current debate in economics thus concerns mainly the role that psychological data and psychological explanations should play in choice theory. Advocates of the 'psychologization' of economics have urged their colleagues to revise the Neoclassical theory of choice, to study the psychological mechanisms underlying decision-making in more depth, and to make extensive use of behavioural and neuroscientific experiments. According to Matthew Rabin, a prominent behavioural economist, for example, 'economists should aspire to make our assumptions about humans as psychologically realistic as possible' (Rabin 2002: 657). This programme is based on a scientific gamble: that economic theory (and choice theory in particular) can be substantially improved by studying the psychology of human decision. The gamble will pay off if there are indeed robust and systematic mechanisms that help understand and predict behaviour more accurately than the standard theory does.

But what does this imply for preferences and beliefs? How are they integrated in the new explanatory framework of behavioural economics? Has a new, psychological concept of preference emerged from the psychology of decision-making? Or has the concept of preference remained unscathed? In the latter case, preferences would continue to be used pretty much as economists have always done.

This is a scientific, rather than a definitional matter. To settle it, we must understand how preferences are typically used by choice theorists, and then see if they are used differently in behavioural economics. The next two sections are devoted to this task.

4. Explaining with preferences

Preferences provide information about the relative attractiveness of different states of affairs. They explain, for example, why Tony usually dines at Pizza Vesuvio rather than Bella Napoli, but also why he would change his habits if Vesuvio were to charge five extra dollars for his favourite pizza. In general, preferences are *explanatory relevant* and help formulate counterfactual claims about future or hypothetical scenarios, which may inform the decisions of scientists and policymakers.[13] Of course preference-based explanations are not the only game in town, and compete with alternative explanations in the scientific arena. Vince's cocaine habit for example escapes choice-theoretic analysis, because it is insensitive to variations in the costs of consumption.[14]

[13]On explanatory relevance and counterfactual reasoning see e.g. Woodward (2004).

[14]I am simplifying here: strictly speaking this would be true only if Vince's demand for cocaine were completely inelastic, which is rarely the case even for drug addicts (Hyman 2009). Many critics of traditional choice theory focus on cases that are conceptually analogous to this: the main effect of moral norms and

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

Even when they are explanatory, however, preferences do not provide information about many interesting questions. They do not tell us, for example, *how* – through which causal mechanism – a given price variation may affect Tony's behaviour. They tell us that A (an agent with certain preferences and beliefs) does B (engages in a certain behaviour) in C (a set of circumstances), without saying exactly how B and C are causally related. Adopting a venerable terminology, throughout the paper I will say that when a system A has a set of causal properties such that, when circumstances C occur, A does B, then *A has a disposition to do B* (in C).[15] I will take this as a stipulative definition, rather than an analysis of the everyday concept of disposition. The point is not to solve an ancient philosophical dispute about a difficult modal notion, but to borrow a useful terminology that helps articulate the main features of preference-based explanations, and show that they share such features with other scientific explanations.

The set of circumstances C is called the 'stimulus' or 'trigger' of B. For example, to say that crystals of NaCl dissolve if you put them in water is equivalent to saying that salt is soluble, and water triggers the dissolution of salt. Thanks to modern chemistry we know quite a lot about this process: roughly, mixing NaCl with H_2O causes the dissociation of NaCl into ions. But knowledge of dispositions can be informative and explanatory even if we do not know the details of the underlying causal processes. Thus when I say that 'the vase is fragile', I am suggesting that we should handle it with care. But knowing the causal process that may lead to the breaking of the vase into little pieces is not necessary for many pragmatic and scientific purposes.

Dispositional views of preferences are traditionally associated with behaviourism. The behaviourist is attracted to dispositions for epistemic reasons: because the causal basis is partly internal and unobservable, the behaviourist despairs that we will ever know anything about it. If A, B and C only refer to observable events or states, then the epistemic worry disappears. Unfortunately, however, we have seen that the connection between choice and preferences in choice theory is mediated by beliefs. Preferences are 'belief-dependent dispositions' (Hausman 2012: 33), that is, the sort of dispositions that will not help the behaviourist project, because unobservable states are included in C.[16]

commitments, for example (Sen 1977), is to make some behaviours (e.g. honest dealings) inelastic to variations in costs and benefits (e.g. bribes). But again, strictly speaking, such cases fall outside the realm of choice theory only in case of complete inelasticity.

[15]I put C between brackets because the triggering conditions are often implicit in dispositional terms: to say that 'petrol is flammable' simply means that it burns at high temperatures. To simplify I am also ignoring *ceteris paribus* clauses and stochastic causation, which are usually implicit in dispositional claims (petrol burns upon reaching a high temperature, if there is oxygen in the environment, and perhaps only with a high probability). Choi and Fara (2018) review and discuss many of these complications.

[16]There is an old tradition in philosophy, going back to Gilbert Ryle, that identifies mental states with dispositions. Since preferences in choice theory fulfil a similar role as desires in folk psychology, it is not surprising to find dispositional accounts of preferences in the contemporary literature (e.g. Sugden 1991: 762; Dowding 2002; Vanderbeeken and Weber 2002; Hausman 2012; Cozic and Hill 2015. *Functionalists*, as we have seen, also claim that mental states like desires are defined partly in terms of causal relations with other mental states (e.g. Block 1980). But since I intend to argue that preferences are not always or necessarily mental states, I will distinguish dispositional from functionalist accounts of preferences here.

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

Dispositions however do achieve some economy of thought. When it is used to explain, a disposition typically brackets the causal basis that connects C (the trigger) with B (the behaviour). In a single formula:

$$C\ [+\ \text{causal basis}] \rightarrow B.$$

Bracketing the causal basis has advantages and limitations. Starting from the latter, it is obvious that dispositions are unfit to answer certain explanatory questions.[17] As Molière pointed out, it is useless to say that opium has *virtus dormitiva*, if the issue is why opium makes you sleep. When the question concerns the details of the causal process, dispositions do not provide useful information.

However, dispositional explanations have the advantage that they allow to express generalizations when the causal basis is *multiply realisable*. A causal basis is multiply realisable when a certain type of event triggers the same effect across a range of circumstances, even though the causal mechanisms differ from case to case. Since economists are usually interested in the explanation of *aggregate* phenomena, dispositional explanations spare them the trouble of giving extremely complicated and heterogeneous descriptions of individual decision processes (Dowding 2002; Ross 2005). But multiple realization may be troublesome even in the case of individuals: for example, suppose that Tony once felt sick after eating pepperoni pizza. Since then, he has refused to eat it in three separate occasions: the first time, Tony recalled the effects of food poisoning and was driven instinctively to avoid pepperoni pizza, following a gut feeling. The second time, Tony's stomach did not react pre-emptively, but Tony deliberated not to eat pepperoni pizza because he remembered the food poisoning. Finally, by the third occasion, Tony had developed a habit to avoid pepperoni pizza without even thinking about it. In spite of the different causal mechanisms, it is correct to generalize across these cases and say that Tony has a belief-dependent disposition (a preference) for not eating pepperoni pizza.

In order to support useful generalizations, dispositions must have some stable features. And in fact most debates in behavioural economics concern the properties that preferences have. So, even though knowledge of dispositions does not require knowledge of their causal basis, understanding the causal basis usually helps modelling the dispositions correctly. The more information we have about the causal mechanisms that are triggered by C, the better we should be able to predict and explain the occurrence of B across various situations.

We can thus distinguish between two explanatory tasks: at one level, economists are interested in providing *explanations of behaviour*; at another level, many (but not all) economists believe that improving such explanations requires that we are also able to *explain preferences*. Another way to put it is to say that preferences are part of the *explanans* (what does the explaining) at the former level, but constitute the *explanandum* (what is to be explained) at the latter. Thus for example *preferences explain behaviour in choice theory;* but *psychological mechanisms explain preferences in behavioural economics*. Or, more precisely, psychological mechanisms explain why certain preferences have the shape that they do have in the models of (behavioural) choice theory.

[17]For a detailed discussion see for example Mumford (1998: ch. 6).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

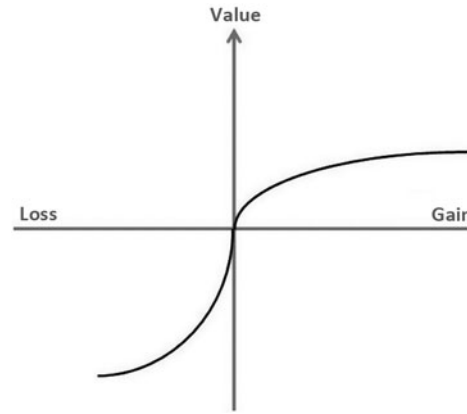


Figure 1. Reference-dependent utility in Prospect Theory.

5. For example: Prospect Theory

The best way to illustrate the difference between explanations of behaviour and explanations of preferences is to examine a specific model, and no model is better suited for this purpose than Prospect Theory. Prospect Theory was developed by Daniel Kahneman and Amos Tversky (1979; see also Tversky and Kahneman 1992) with the aim of systematizing a large body of experimental results produced by psychologists of decision-making during the 1960s and 70s. The two main features of Prospect Theory – which distinguish it from traditional models of rational choice – are the distortion of subjective probabilities (beliefs) and the reference-dependent utility function. Since preferences are the main topic of this paper, I will ignore beliefs and focus on the latter only.

The classic shape of Kahneman and Tversky's utility function is represented in Figure 1. The intersection of the horizontal and vertical axes denotes the reference point, that is, the state of affairs with respect to which the agent evaluates the possible outcomes of her decision. In many cases the reference point is simply the status quo. The first important idea behind Prospect Theory is that the same outcome may be evaluated differently by the same individual depending on whether it is perceived as a gain or a loss with respect to the reference point. The second idea is that utility decreases in the domain of losses more steeply than it increases in the domain of gains. The 'kink' in the shape of the utility function reflects this asymmetry in the evaluation of gains and losses.

Although at times they seem to differentiate their language on purpose – they speak of 'value functions' instead of 'utility functions', for example – in most respects Kahneman and Tversky do their best to present Prospect Theory in the same format as traditional choice theory.[18] Its psychological content lies in the mechanisms that determine the evaluation of prospects, but none of this implies that *the value function* must be interpreted psychologically. The preferences of Prospect Theory in fact do not look different from those of traditional choice theory – they describe belief-dependent

[18]The original paper included an appendix with a representation theorem, for example. See also Wakker (2010) for an attempt to integrate Prospect Theory in the traditional conceptual apparatus of economic theory.

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

dispositions. The principal difference is that Prospect Theory uses psychological information to attain a more accurate representation of preferences, whereas traditional choice theory tends to sacrifice realism in favour of normative considerations (its agents are *rational* decision-makers, first and foremost).[19]

What is the causal basis of preferences according to Prospect Theory? Terms like 'loss aversion' and 'reference dependence' refer to systematic biases in the evaluation of outcomes.[20] So by looking at the utility function we cannot tell what kind of psychological processes lie behind Prospect Theory. Kahneman and Tversky however offer an informal story of how the evaluation of prospects takes place: in the so-called *editing phase* of decision, the decision-maker engages in a 'preliminary analysis of the offered prospects, which often yields a simpler representation of these prospects' (1979: 274). The peculiar shapes of the probability and value functions are generated at this stage by operations such as 'coding', 'combination', 'segregation' and 'cancellation', which are meant to facilitate the representation and evaluation of the prospects.

This is roughly the psychological content of Prospect Theory. Although the later literature has made other contributions to the explanation of utility functions, such contributions tend to fall in the area of evolutionary psychology or neuroscience, and as such they do not concern the proximate mechanisms of decision. Reference dependence for example may be explained by the fact that our cognitive system is essentially a detector of novelties.[21] New events convey potentially important information regarding opportunities and risks, to which our cognitive apparatus must often react quickly. One cheap and dirty way to do it is to assess their positive or negative valence with respect to the status quo, so as to determine promptly whether changing one's behaviour is worthwhile. Loss aversion – the fact that the utility function is steeper in the domain of losses – may be explained by the fact that losses and gains are processed in different parts of the brain (Blackford *et al.* 2010). Losses tend to trigger negative emotional reactions, and the 'loss module' may be over-sensitive for evolutionary reasons: for a creature living in a dangerous environment and always on the brink of extinction, trying to avoid fatal losses may be more sensible (in fitness terms) than seeking uncertain gains.[22]

Although such explanations support the story told by Kahneman and Tversky, they are one step behind the psychological mechanisms studied by behavioural economists. They offer evolutionary or neuroscientific reasons to believe in the existence of such mechanisms, without shedding further light on the processes leading to decisions. But above all, such research does not bear on the interpretation of preferences. Nothing in cognitive or neuroscience forces a mentalistic interpretation across the board: although Kahneman and Tversky explain why *individual human* preferences have a particular shape, their theory is compatible with the idea that preferences are variably realized across different decision units. Moreover, the causal

[19]On the importance of normative principles in choice theory, see Sugden (1991), Guala (2000), Starmer (2000), Okasha (2016).

[20]Some critics, in fact, have pointed out that 'several of the explanations using heuristics are hardly more than redescriptions of the phenomena' (Gigerenzer 1991: 99).

[21]See e.g. Tiitinen *et al.* (1994), Corbetta and Shulman (2002).

[22]A piece of evidence in favour of this hypothesis, for example, is that loss aversion generally increases when the budget is small (Camerer 2005).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

basis of some preferences, in some collective organizations and non-human organisms, may not be psychological in any interesting sense of the term. In the next section we will examine two decision units that display similar preference patterns for completely different reasons. But only in one case, as we shall see, is the causal basis psychological in an interesting sense of the term.

6. When the causal basis varies: intransitive preferences

The rejection of mentalism is based on a distinction between different levels of explanation. And the distinction between levels is justified by multiple realizability: if the causal basis of preferences differs across decision units, different types of explanation may be in order for different domains of application of choice theory. But multiple realizability, of course, is an empirical matter: the causal bases of preferences may or may not be psychological, in each particular case. Tony's dislike of pepperoni pizza was prompted in one case by an emotional reaction, another time by deliberate reasoning, and the third time by an ingrained habit that, presumably, must be encoded in his brain.[23] In this case, there does not seem to be any reason to deny that preferences are psychological dispositions. But in other cases, the same conclusion is not warranted at all.

I will proceed again by way of an example. Experimental and behavioural economists have demonstrated convincingly that preferences may be intransitive in some circumstances, and have described various mechanisms that may explain these intransitive patterns. One mechanism was identified a long time ago by Tversky and other choice theorists working on so-called 'multi-attribute' utility functions. Imagine a decision-maker facing a choice between three options labelled X, Y and Z. Each option moreover is characterized by three attributes. The options and the attributes may be anything that the decision-maker cares about, but to make the example concrete let us suppose that the choice concerns three types of car characterized by different degrees of comfort, speed and price. Now, let us suppose that the cars (X, Y, Z) are ranked on each dimension or characteristic as follows:

1. Comfort: X>Y>Z.
2. Speed: Z>X>Y.
3. Price: Y>Z>X.

The > symbol stands for a (transitive) 'more-than' relation: for example, along the first dimension car X is more comfortable than Y, which is more comfortable than Z. Now, since no car is better than the others in all respects, the decision-maker faces the problem of weighing each attribute and aggregating the evaluations to form a single preference ranking. This task may be cognitively quite demanding, so when presented with pairwise choices people tend to use a simple heuristic instead of working through the aggregation procedure. The heuristic goes as follows: whenever you have to make a decision among two options, choose the one that

[23]Some contemporary attempts to identify 'utility in the brain' are probably based on this sort of reasoning. If not in the mind/brain, where could utility possibly be? See e.g. Glimcher *et al.* (2005), and – for a critical perspective – Vromen (2010), Fumagalli (2013).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

beats the other along most dimensions. Applied to this particular case, the rule would give the following outcomes:

X beats Y in comfort and speed, but loses in price.
Y beats Z in comfort and price, but loses in speed.
Z beats X in speed and price, but loses in comfort.

Using this simple heuristic would generate intransitive preferences: X is preferred to Y, Y is preferred to Z, but Z is preferred to X.[24]

The mechanism behind preference formation, in this example, is clearly psychological: the main reason to use the simple rule is to save cognitive effort. Each one of us can appreciate the lure of this heuristic by introspection, and experiments (e.g. Tversky 1972) have demonstrated that human decision-makers have a tendency to engage in mono-dimensional evaluation before (or instead of) aggregating across the attributes. A mentalistic interpretation of the causal basis of preferences in this case is empirically and theoretically warranted.

It is easy however to find examples of intransitive preferences that are not based on a psychological mechanism of this sort. For ease of comparison, let us examine a case that is formally similar to the previous one: a university hiring committee must decide how to rank three candidates (Dr Smith, Dr Brown and Dr Jones) for an academic position. The committee is constituted by three members, who rank the candidates as follows:

1. Smith > Jones > Brown.
2. Brown > Smith > Jones.
3. Jones > Brown > Smith.

Now, let us suppose that according to university rules the winner must be determined via a selection process that involves a sequence of pairwise comparisons, and that each comparison must be resolved by majority voting. When presented with pairwise options, the committee votes as follows:

Smith beats Jones 2-1.
Jones beats Brown 2-1.
Brown beats Smith 2-1.

This is a classic example of 'Condorcet voting', a problem that has been discussed extensively in choice theory.[25] Condorcet voting is a case of intransitivity generated by voting rules. A non-psychological mechanism here produces a peculiar pattern of preferences. Although formally analogous, the mechanism thus is substantially different from the one that explains intransitivity in multi-attribute choice. The first difference is that the decision-maker is an organization rather than an individual human being. The second one is that intransitive preferences are generated by

[24]For a classic experimental study of this phenomenon, see May (1954). Other excellent discussions of intransitivity can be found in Tversky (1969) and Mongin (2000b).

[25]See e.g. Farquharson (1969), List and Pettit (2011).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

an institutional rule, rather than by a heuristic.[26] Any attempt to fix the problem, therefore, should better take this fact into account: the committee does not suffer from bounded rationality, does not follow fast-and-frugal heuristics, and transitivity cannot be restored by, say, changing the way in which the options are framed. An effective intervention must target the institutional environment, for example the rules that force the committee to follow a pairwise, majority-voting procedure.

It is important to realize that this is not an odd application of choice theory. Although the theory was originally developed to explain the behaviour of individual human beings, economists and other behavioural scientists use it routinely to explain and predict the choices of other decision units. We know that mice and crabs have preferences – in the genuine sense of the term – because we can find their indifference points by varying costs and benefits and observing that their behaviour changes in a systematic manner. If laboratory mice are made to work to obtain different types of water, for example, their behaviour fits a well-defined demand curve (Holm *et al.* 2007). If hermit crabs are given electric shocks, they are more reluctant to leave a good shell than a poor quality one (Elwood and Appel 2009). Even jellyfish, according to marine biologists, have a control centre that 'coordinates responding when stimuli are tending to elicit more than one response' (Albert 2011: 474).

And if these applications seem marginal, consider traditional analyses of consumers' choice: economists are perfectly comfortable to apply choice theory to explain the consumption patterns of households. And they are well aware that the preferences of a household may be irreducible to the preferences of its individual members. Thus, for example, sea and mountain resorts are perfectly complementary in the utility function of Vince's family. This property does not reflect anyone's individual preferences – Vince prefers mountains, Margie prefers sea – and cannot be explained psychologically. Rather, Vince and his family have decided to adopt a simple rule: to make everyone happy, they spend half of their holidays in the mountains and the other half on the beach.

One may worry at this point that we are injecting an odd asymmetry in choice-theoretic explanations: why should we take only one of the partners that explain and justify choices to be mental, and not the other? The worry however seems unjustified: the beliefs of committees, robots and non-human organisms can also be interpreted in a non-mentalistic fashion. For example, the information that is available to a selection committee – in the form of CVs, letters of reference and interviews – may influence its decisions, without being identical to the psychological states of the committee members. There is no reason why the information should be treated (represented, stored, revised) in a way that satisfies the laws of individual human psychology: the weight accorded to a piece of information, for example, may depend on strategic coalition-formation among the committee members or on institutional rules that determine its eligibility, rather

[26]There are, of course, psychological elements in the causal chain that leads to the committee's preference structure (such as the evaluations of the individual committee members). But such psychological elements cannot explain the intransitivity of preferences – not, at least, without a lot of help from the institutional rules (see also Ross 2005, 2014).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

than on its salience, availability or the cognitive limitations of the committee members.[27] To sum up: the causal basis of beliefs, preferences, and the way in which they interact is not always, or not even typically, constituted by psychological processes.

A committed functionalist may still insist that there is no reason to identify mental states narrowly with human mental states, and psychology with human psychology. Following List and Pettit (2011), for example, one may want to extend mentalism to corporate agents. But this move depends on the existence of a *science* of the mental that is able to provide unifying explanations of the preferences of human individuals, social groups, crabs, robots and other decision units that fall in the domain of choice theory. As things stand, it seems unlikely that a single theory will be able to account for the causal basis of preferences across such a wide domain. This is one of the reasons why economists prefer to be non-committal about the nature of preferences. A functionalist mentalism may seem appealing to philosophers, but would be largely irrelevant for scientific practice.

7. Taking stock

The controversy on the interpretation of utility and preferences is one of the oldest debates in the foundations of economics. Its resilience is due to an unresolved tension between, on the one hand, economists' aspiration to theorize at a different level from psychology, and, on the other, the fact that psychological evidence has been an important source of information to build modern microeconomic theory. This has become particularly evident during the last three decades, since the rise of behavioural economics has led to the development of psychologically informed models of choice.

Seen in this light, the resistance against mentalism may appear all the more obsolete. Nonetheless, I have argued that the notion of preference is used in choice theory – including *behavioural* choice theory – in a way that does not compel a psychological interpretation. The reason is that preferences are (*belief-dependent*) *dispositional terms with a multiply realizable causal basis*. Because they are *belief-dependent* dispositions, behaviourism fails. But because they are *multiply realizable*, mentalism does not hold. When the causal basis of preferences is institutional (as in the case of organizations), bio-physical (as in jellyfish), or mechanical (as in robots), standard psychological explanations are unable to shed light on the causal basis and the shape of preferences.

The argument hinges on a disambiguation of the term 'mental': for most economists it denotes the realm of phenomena and processes studied by the psychology of human decision.[28] This is often a cause of misunderstanding, for philosophers tend to use the term in a different, more technical way. Functionalists, for example, would agree with the main premise of the argument (preferences are

[27]While serving as members of hiring committees, most of us have heard the expression 'unfortunately, we cannot consider this information' or 'we cannot include it in the candidate's file'.

[28]Gul and Pesendorfer (2008) adopt a narrower interpretation, suggesting that 'mental' is synonymous with 'neural'. This is a misleading choice of terminology, for it seems to deny the possibility of theorizing at a higher level than brain states and processes – which is what most behavioural economists and economic psychologists do. On this point, see also Dietrich and List (2016: 276).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

dispositions) but would take it to corroborate rather than to refute mentalism. As I have argued in Section 3, this is not merely a terminological quibble. The functionalist interpretation casts a shadow of mystery on the controversy sparked by the mentalistic interpretation. The mystery disappears once we realize that the controversy concerns the relationship between different scientific programmes and explanatory frameworks within the sciences of behaviour. Anti-mentalists in economics avoid the identification of choice theory with psychology in order to distinguish between different explanatory tasks, and to make its application possible in a wider domain. This is the main reason why they remain non-committal about the causal basis of preferences.

In some circumstances, especially in the traditional domain of human decision-making, the causal basis is indeed mainly constituted by psychological mechanisms. The dispositional interpretation in such cases clarifies how economists and cognitive scientists collaborate, indicating a natural division of labour and providing a common language for the construction of psychologically informed models of choice.[29] But in other cases the basis is not psychological in any interesting, scientific sense of the term. This explains why some economists resist the narrow identification of choice theory with the psychology of decision-making.

The latter is not an unscientific or a conservative position. Like 'force' in physics, 'preference' is an abstract concept that must be filled with content in each specific domain of application. Like physical forces, we have good reasons to believe that preferences are real explanatory variables. But what they are, exactly, is something that science – not philosophy – will tell. To discover how many kinds of force there are, and how they work, has been an important task in the history of physics. Similarly, the mechanisms underlying preferences are potentially heterogeneous and new ones may be discovered with the development of social science.

This is not a reductionist position either, from a metaphysical or methodological point of view. One may refuse to identify preferences with specific psychological states and at the same time recognize that they are genuine (i.e. causally effective) explanatory variables.[30] The point is that as dispositions they may have different properties depending on the type of decision unit: their causal basis is likely to differ across domains. A deeper explanation, as mentioned in Section 4, is not always necessary or even useful for many practical and scientific purposes. But any attempt to explain preferences (why the preferences of a university committee have a certain shape, for example, while those of a mobster have another) will most likely draw upon theoretical and methodological resources that are suited to the specific causal basis of each domain. 'Sociology for committees, engineering for robots and psychology for humans', may be the slogan of this paper.

Acknowledgements. The ideas that form the core of this paper were presented at seminars held at the universities of Amsterdam, Bayreuth, Florence, Innsbruck, Helsinki, LSE, Paris 1, Radboud, Rotterdam and Turin. I'm grateful to Constanze Binder, Mikael Cozic, Franz Dietrich, Roberto Fumagalli, James Grayot, Cyril Hédoin, Conrad Heilmann, Christian List, Caterina Marchionni, Michiru Nagatsu, Gianluca Pozzoni, Bob Sugden, Kate Vredenburgh, Jack Vromen, Bernard Walliser, Philip Wichardt

[29]Arguably, behavioural economics is committed not only to study the psychological basis of economic behaviour, but also to translate its findings in the language of economic theory.

[30]On non-reductive materialism and causes as difference-makers, see List and Menzies (2009).

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

and two anonymous referees for their suggestions, comments and critiques. Ivan Moscati deserves special gratitude for several conversations that motivated and shaped the writing of the paper. All the mistakes are, of course, my exclusive responsibility.

References

Afriat S. N. 1967. The construction of utility functions from expenditure data. *International Economic Review* **8**, 67–77.

Albert D. J. 2011. What's on the mind of a jellyfish? *Neuroscience and Biobehavioral Reviews* **35**, 474–482.

Angner E. 2018. What are preferences? *Philosophy of Science* **85**, 660–681.

Binmore K. 1994. *Game Theory and the Social Contract, Vol. 1: Playing Fair*. Cambridge, MA: MIT Press.

Binmore K. 2008. *Rational Decisions*. Princeton, NJ: Princeton University Press.

Blackford J. U., J. W. Buckholtz, S. N. Avery and D. H. Zald 2010. A unique role for the human amygdala in novelty detection. *Neuroimage* **50**, 1188–1193.

Block N. 1980. Troubles with functionalism. In N. Block (ed.), *Readings in Philosophy of Psychology*. Vol. **1**, Cambridge, MA: Harvard University Press, pp. 261–325.

Bridgman P. W. 1927. *The Logic of Modern Physics*. New York, NY: Arno Press.

Bruni L. and R. Sugden 2007. The road not taken: how psychology was removed from economics, and how it might be brought back. *Economic Journal* **117**, 146–173.

Camerer C. 2005. Three cheers – psychological, theoretical, empirical – for loss aversion. *Journal of Marketing Research* **42**, 129–133.

Camerer C. 2008. The case for mindful economics. In A. Caplin and A. Schotter (eds), *The Foundations of Positive and Normative Economics*. New York, NY: Oxford University Press, pp. 43–69.

Caplin A. 2008. Economic theory and psychological data: bridging the divide. In A. Caplin and A. Schotter (eds), *The Foundations of Positive and Normative Economics*. New York, NY: Oxford University Press, pp. 336–371.

Choi S. and M. Fara 2018. Dispositions. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*. Available at http://plato.stanford.edu/entries/dispositions.

Clarke C. 2016. Preferences and positivist methodology in economics. *Philosophy of Science* **83**, 192–212.

Corbetta M. and G. L. Shulman 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* **3**, 201–215.

Cozic, M. and B. Hill 2015. Representation theorems and the semantics of decision-theoretic concepts. *Journal of Economic Methodology* **22**, 292–311.

Dietrich F. and C. List 2016. Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics and Philosophy* **32**, 249–281.

Dowding K. 2002. Revealed preference and external reference. *Rationality and Society* **14**, 259–284.

Edwards J. 2016. Behaviorism and control in the history of economics and psychology. *History of Political Economy* **48**, 170–197.

Elwood R. W. and M. Appel 2009. Pain experience in hermit crabs? *Animal Behaviour* **77**, 1243–1246.

Engelen B. 2017. A new definition of and role for preferences in positive economics. *Journal of Economic Methodology* **24**, 254–273.

Farquharson R. 1969. *Theory of Voting*. Oxford: Blackwell.

Fumagalli R. 2013. The futile search for true utility. *Economics and Philosophy* **29**, 325–347.

Gigerenzer G. 1991. How to make cognitive illusions disappear: beyond "heuristics and biases". *European Review of Social Psychology* **2**, 83–115.

Glimcher P. W., M. C. Dorris and H. M. Bayer 2005. Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior* **52**, 213–256.

Guala F. 2000. The logic of normative falsification: rationality and experiments in decision theory. *Journal of Economic Methodology* **7**, 59–93.

Guala F. 2012. Are preferences for real? Choice theory, folk psychology, and the hard case for commonsensible realism. In A. Lehtinen and P. Ylikoski (eds), *Economics for Real*. London: Routledge, pp. 137–155.

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

Gul F. and W. Pesendorfer 2008. The case for mindless economics. In A. Caplin and A. Schotter (eds), *The Foundations of Positive and Normative Economics*. New York, NY: Oxford University Press, pp. 3–39.

Hands D. W. 2009. Economics, psychology and the history of consumer choice theory. *Cambridge Journal of Economics* **34**, 633–648.

Hands D. W. 2013. Foundations of contemporary revealed preference theory. *Erkenntnis* **78**, 1081–1108.

Hands D. W. 2014. Paul Samuelson and revealed preference theory. *History of Political Economy* **46**, 85–116.

Hausman D. M. 2000. Revealed preference, belief, and game theory. *Economics and Philosophy* **16**, 99–115.

Hausman D. M. 2008. Mindless or mindful economics: a methodological evaluation. In A. Caplin and A. Schotter (eds), *The Foundations of Positive and Normative Economics*. New York, NY: Oxford University Press, pp. 125–151.

Hausman D. M. 2012. *Preference, Value, Choice and Welfare*. New York, NY: Cambridge University Press.

Hicks J. 1939. *Value and Capital*. Oxford: Oxford University Press.

Holm L., C. Ritz and J. Ladewig 2007. Measuring animal preferences: shape of double demand curves and the effect of procedure used for varying workloads on their cross-point. *Applied Animal Behaviour Science* **107**, 133–146.

Houthakker H. 1950. Revealed preference and the utility function. *Economica* **17**, 159–174.

Hyman G. M. 2009. *Addiction: A Disorder of Choice*. Princeton, NJ: Princeton University Press.

Kahneman D. and A. Tversky 1979. Prospect Theory: an analysis of decision under risk. *Econometrica* **47**, 263–291.

Kahneman D., P. P. Wakker and R. Sarin 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* **112**, 375–405.

Lehtinen A. 2011. The revealed preference interpretation of payoffs in game theory. *Homo Oeconomicus* **28**, 265–296.

Lewin S. 1996. Economics and psychology: lessons for our own day from the early twentieth century. *Journal of Economic Literature* **34**, 1293–1323.

List C. and P. Menzies 2009. Nonreductive physicalism and the limits of the exclusion principle. *Journal of Philosophy* **106**, 475–502.

List C. and P. Pettit 2011. *Group Agency*. New York, NY: Oxford University Press.

Mas-Colell A., M. D. Whinston and J. R. Green 1995. *Microeconomic Theory*. Vol. **1**. New York, NY: Oxford University Press.

May K. O. 1954. Intransitivity, utility, and the aggregation of preference patterns. *Econometrica* **22**, 1–13.

Mongin P. 2000a. Les préférences révélées et la formation de la théorie du consommateur. *Revue économique* **51**, 1125–1152.

Mongin P. 2000b. Does optimization imply rationality? *Synthese* **124**, 73–111.

Moore J. 2001. On distinguishing methodological from radical behaviorism. *European Journal of Behavior Analysis* **2**, 221–244.

Moscati I. 2018. *Measuring Utility. From the Marginal Revolution to Behavioral Economics*. New York, NY: Oxford University Press.

Moscati I. (Forthcoming). Not a behaviourist: Samuelson's contributions to utility theory in the Harvard years, 1936–1940. In R. Anderson, W. Barnett and R. A. Cord (eds), *Paul Samuelson: Master of Modern Economics*. London: Palgrave Macmillan.

Mumford S. 1998. *Dispositions*. Oxford: Oxford University Press.

Nagatsu M. and K. Poder (Forthcoming). Economists' concept of choice: an experimental study. *Economics and Philosophy*.

Okasha S. 2016. On the interpretation of decision theory. *Economics and Philosophy* **32**, 409–433.

Pareto V. 1906. *Manuale d'economia politica*. Milano: Società Editrice Libraria.

Putnam H. 1975. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.

Rabin M. 2002. A perspective on psychology and economics. *European Economic Review* **46**, 657–685.

Rosenberg A. 1993. *Economics: Mathematical Politics or Science of Diminishing Returns?* Chicago, IL: University of Chicago Press.

Ross D. 2005. *Economic Theory and Cognitive Science*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press

Ross D. 2011. Estranged parents and a schizophrenic child: choice in economics, psychology and neuroeconomics. *Journal of Economic Methodology* **18**, 217–231.
Ross D. 2014. Economic versus psychological models of bounded rationality. *Journal of Economic Methodology* **21**, 411–427.
Samuelson P. A. 1938. A note on the pure theory of consumer's behaviour. *Economica* **5**, 61–71.
Samuelson P. A. 1948. Consumption theory in terms of revealed preferences. *Economica* **15**, 243–53.
Samuelson P. A. 1950. The problem of integrability in utility theory. *Economica* **17**, 355–383.
Sen A. K. 1977. Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs* **6**, 317–344.
Sen A. K. 1982. *Choice, Welfare and Measurement*. Oxford: Blackwell.
Starmer C. 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* **38**, 332–382.
Sugden R. 1991. Rational choice: a survey of contributions from economics and philosophy. *Economic Journal* **101**, 751–785.
Tiitinen H., P. May, K. Reinikainen and R. Näätänen 1994. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* **372**, 90–92.
Tversky A. 1969. Intransitivity of preferences. *Psychological Review* **76**, 31–48.
Tversky A. 1972. Elimination by aspects: a theory of choice. *Psychological Review* **79**, 281–299.
Tversky A. and D. Kahneman 1992. Advances in Prospect Theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5**, 297–323.
Vanderbeeken R. and E. Weber 2002. Dispositional explanations of behavior. *Behavior and Philosophy* **30**, 43–59.
Varian H. 2005. *Intermediate Microeconomics*. London: Norton.
Vromen J. 2010. On the surprising finding that expected utility is literally computed in the brain. *Journal of Economic Methodology* **17**, 17–36.
Wakker P. 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press.
Watson J. B. and W. McDougall 1929. *The Battle of Behaviorism*. New York, NY: Norton.
Woodward J. 2004. *Making Things Happen*. New York, NY: Oxford University Press.

Francesco Guala is Professor of Political Economy in the Department of Philosophy of the University of Milan. A believer in interdisciplinarity, he has published widely in philosophy, economics and psychology journals. His monographs include *The Methodology of Experimental Economics* (Cambridge University Press, 2005), and *Understanding Institutions* (Princeton University Press, 2016). He has co-edited, with Daniel Steel, *The Philosophy of Social Science Reader* (Routledge, 2011).

Cite this article: Guala F (2019). Preferences: neither behavioural nor mental. *Economics and Philosophy* **35**, 383–401. https://doi.org/10.1017/S0266267118000512

https://doi.org/10.1017/S0266267118000512 Published online by Cambridge University Press