

PREFERENCES: NEITHER BEHAVIOURAL NOR MENTAL

FRANCESCO GUALA

Working Paper no. 5/2017

FEBRUARY 2017

 UNIVERSITÀ DEGLI STUDI DI MILANO



DIPARTIMENTO DI ECONOMIA, MANAGEMENT E METODI QUANTITATIVI

Via Conservatorio 7
20122 Milano

tel. ++39 02 503 21501 (21522) - fax ++39 02 503 21450 (21505)

<http://www.economia.unimi.it>

E Mail: dipeco@unimi.it

Preferences: Neither Behavioural nor Mental*

Francesco Guala

University of Milan, Italy

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 dispositions with a multiply realizable causal basis, which explains why economists are reluctant to make a commitment about their interpretation.

We *don't* say 'solubility is a certain physical-chemical structure', but rather that the solubility of those substances that are soluble is *explained* by their possession of a certain physical-chemical structure. Similarly [...] if we say [...] that preferring *A* to *B* is 'synthetically identical with' possessing a certain physical-chemical structure [...] then we let ourselves in for what seems to me remarkable and insufficiently motivated extensions of usage. (Putnam 1975: 417)

1. Introduction

Some philosophical controversies never die. The debate about the interpretation of preferences and utility¹ in economics flared up at the end of the nineteenth 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

* The ideas that form the core of this paper were presented at seminars held at the universities of Bayreuth, Innsbruck, Helsinki, Paris 1, and Turin. I'm grateful to Mikael Cozic, Franz Dietrich, Roberto Fumagalli, Cyril Hedoin, Christian List, Caterina Marchionni, Gianluca Pozzoni, Bob Sugden, Jack Vromen, Philip Wichardt, and many other participants to the above seminars for their suggestions, comments and critiques. Ivan Moscati deserves special gratitude, for several conversations that motivated the writing of this paper. All the mistakes are, of course, my exclusive responsibility.

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 For a historical reconstruction of some key aspects of the old controversy, see Lewin (1996), Bruni and Sugden (2007), Moscati (2013a, 2013b, 2015).

3 An incomplete list of contributions to the recent debate includes Hausman (2000, 2008, 2012), Mongin (2000a), Dowding (2002), Glimcher et al (2005), Ross (2005, 2011, 2014), Camerer (2008), Gul and Pesendorfer (2008),

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 utility 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 not mental states: they 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, however. On the contrary it is entirely compatible with the idea that psychology plays an important role in choice theory. But this does not imply that preferences or utility 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 traditional domain of human decision making. This extension requires that ‘preference’ is given an abstract interpretation, along the same lines as concepts like ‘force’ in physics or ‘fitness’ in biology. 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 paper is organized as follows: section two 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 three explains what role preferences play in economic explanations of behaviour, in (broadly) dispositional terms. Section four shows that this interpretation is consistent with scientific practice in behavioural economics, focusing on the case of Prospect Theory. In section five I illustrate how intransitive preferences can be multiply realized, using examples from human decision making and committee deliberation. Section six summarizes the argument and concludes the main body of the paper. The last two sections are optional, and written mainly for the sceptics. Section seven defuses a common worry, namely that the proposed interpretation of preferences makes the economic theory of choice void of any substantial content. Section eight criticizes the attempt to defend mentalism from a functionalist perspective.

Hands (2009, 2013, 2014), Vromen (2010), Lehtinen (2011), Guala (2012), Fumagalli (2013), Clarke (2016), Dietrich and List (2016), Okasha (2016), Angner (unpublished), Cozic (unpublished), Nagatsu and Poder (unpublished).

2. Background: why preferences cannot be just behaviour

The interpretation of preferences and utility that became orthodox in the middle of the twentieth 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 concept in terms of measurable behaviour (e.g. Watson and McDougall 1929). Terms like 'hunger' or 'pain', should 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, and Hicks and Allen 1934) 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 and utility.

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 Hicks and Allen's 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). Whereas in Hicks and Allen's approach the concept of preference is primitive, Samuelson took 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 even mentioning utility or preferences, the same results that had been achieved earlier by Pareto, Hicks and Allen.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

4 Although the 'behaviourist' interpretation of Samuelson's project is majoritarian, it has its critics. Mongin (2000a) for example argues that Samuelson never intended the theory of revealed preferences to replace ordinal utility analysis. For a detailed discussion of the relationship between revealed preference and ordinal utility theory, see Cozic (unpublished).

5 See e.g. Samuelson (1948), Houthakker (1950), Afriat (1967). Hands (2013) proposes a useful taxonomy of

(e.g. Mas-Colell et al 1995, Varian 2005). But in spite of its apparent success, the original positivist 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 *ordinal* utility theory, and hence worked within the domain of riskless decisions. 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.
- (b) Vincent prefers the restaurant Pizza Bella Napoli to Pizza Vesuvio. He believes that both are open. Therefore, he goes to Pizza Bella Napoli.

Although Tony and Vincent's 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 else 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 contribute 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 or not, 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. But 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 sophisticated interpretations of choice theory that attempt to circumvent the problem in ingenious ways, but none of

different versions of the revealed preference project.

⁶ For a comprehensive reconstruction of the evolution of Samuelson's position, see Hands (2014).

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 defined in non-behavioural terms), 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 of preferences 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 implied by Samuelson’s choice of terminology,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

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 ‘mentalist’ interpretation of preferences?

3. Explaining with preferences

Most contemporary economists do not care much about behaviourism. 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 start from a common – and seemingly plausible – presumption, namely, that the failure of behaviourism would open the doors to mentalism in economics. The main thesis of this paper is that this worry is unjustified: there is room for a third position, according to which preferences are neither behavioural nor mental.

I will argue later that the third position is consistent with best scientific practice. It is, in particular, consistent with the way in which the concepts of preference and utility are used in behavioural economics – a scientific programme that is entirely devoted to promote the integration of psychology with economics. I will argue that although behavioural economics has undoubtedly introduced psychological concepts and mechanisms in economic science, it has given no reason to interpret utility or preferences as mental states. Preferences, as we shall see, are interpreted in behavioural economics pretty much in the same way as in

7 See Hausman (2000, 2012), Hands (2013) for discussion.

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 firmly entrenched in the literature: see e.g. Hausman (2000, 2012), Hands (2013), Dietrich and List (2016), and, for behaviourism more generally, also Moore (2001).

10 Gul and Pesendorfer (2009) 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, section 4) and Okasha (2016). Sen (1982) is the precursor of most contemporary discussions on the use of preferences in normative economics.

traditional (rational) choice theory. As a preliminary step, therefore, it is necessary to explain how preferences are typically used by economists to explain choice.

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 (at Vesuvio they sell a tastier, cheaper Margherita pizza), but also why he would change his habits if Vesuvio charged 5 extra dollars for his favourite pizza. In general, preferences are *explanatory relevant* and help formulating counterfactual claims about future or hypothetical scenarios, which may inform the decisions of scientists and policy-makers.¹¹ At the same time, preference-based explanations are not trivial and compete with alternative explanations in the scientific arena. Tony's cocaine habit for example escapes choice-theoretic analysis, because it is insensitive to variations in the costs of consumption.¹²

Even when they are genuinely 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) does B (engages in a certain behaviour) in C (a set of circumstances), without saying how B and C are causally related. Adopting a venerable terminology, throughout the paper I will assume that to say that 'A does B in circumstances C' is equivalent to saying that A has a *disposition* to do B.¹³ 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 'trigger' of B. For example, to say that crystals of NaCl dissolve if you put them in water is equivalent to say 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 H₂O causes the dissociation of NaCl into ions. But knowledge of dispositions is informative and explanatory even if we do not know the details of the underlying causal processes. Thus if I say that 'the vase is fragile', I am suggesting that we should handle it with care. But knowledge of the causal mechanism that may lead to the breaking of the vase into little pieces is not necessary for many pragmatic 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

11 On explanatory relevance and counterfactual reasoning see e.g. Woodward (2008).

12 I am simplifying here: strictly speaking this is true only if Vince's demand cocaine is completely inelastic, which is rarely true even for hard 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 commitments, for example (Sen 1977), is to make specific 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.

13 I do not pretend to be particularly original: 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; Vanderbeeken and Weber 2002; Dowding 2002; Hausman 2012; Cozic, unpublished).

unobservable, the behaviourist despairs that we will ever know anything about it. If 'A does B in C' only refers to observable events or states, 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. To avoid confusion, philosophers tend to use a completely different label – 'functionalism' – for the idea 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* mental states, I will avoid that terminology here.¹⁴

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.¹⁵ As Molière pointed out, it is useless to say that opium has *virtuus 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 realizable*.

A causal basis is multiply realizable 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 us the trouble of giving extremely complicated and heterogeneous descriptions of the causes of behaviour (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 decided instinctively not to eat pepperoni pizza, following his 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 routine to avoid pepperoni pizza without even thinking about it. In spite of the different causal mechanisms, in all these cases it is correct to say that Tony has a disposition (a preference) for not eating pepperoni pizza.

In order to support scientific 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

14 I will return to functionalism and mentalism in the last section of this paper.

15 For a detailed discussion see e.g. Mumford (1998, chapter 6).

may help to model the dispositions correctly. The more information we have about the causal mechanisms that connect C with B, 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 giving 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 *explananda* (what is to be explained) at the latter. Thus *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 choice theory.

4. 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 task than Prospect Theory. Prospect Theory was developed by Daniel Kahneman and Amos Tversky (1979; see also Tversky & 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.

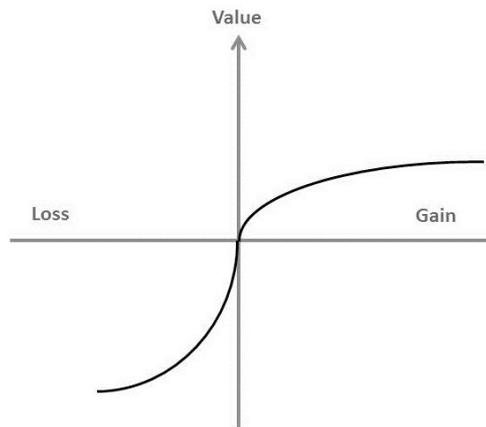


Figure 1: Reference-dependent utility in Prospect Theory.

Notice that so far there is no reason to interpret the preferences summarized in Kahneman and Tversky's utility function as psychological states. 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 theories.¹⁶ All the psychological content lies in the experimental phenomena – like loss aversion – that Kahneman and Tversky rely upon to justify the unorthodox shape of their utility function. But this does not imply that the function must be interpreted psychologically. The preferences of Prospect Theory are not different from those of traditional choice theory – they describe behavioural 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 (the agents of the traditional theory are *rational* decision-makers, first and foremost).¹⁷

What is the causal basis of preferences according to Prospect Theory? 'Loss aversion' or 'reference dependence' refer to experimental effects, rather than causal mechanisms. So it is not clear what kind of psychological processes lie behind Prospect Theory. Although the theory itself is remarkably silent about this aspect, the cognitive science literature offers many useful suggestions. Reference dependence for example may be explained by the fact that our cognitive system is essentially a detector of novelties.¹⁸ 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 would improve or worsen one's situation.

Loss aversion – the fact that the utility function is steeper in the domain of losses – may be

¹⁶ 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.

¹⁷ On the importance of normative principles in choice theory, see Sugden (1991), Guala (2000), Starmer (2000), Okasha (2016).

¹⁸ See e.g. Tiitinen et al. (1994), Corbetta and Shulman (2002).

explained by the fact that losses and gains are processed in different parts of the brain.¹⁹ 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 is more sensible (in fitness terms) than seeking uncertain gains. A piece of evidence in favour of this hypothesis, for example, is that loss aversion generally increases when the budget is small (Camerer 2005).

Prospect Theory, thus, is not a theory of psychological preferences – it is a *psychological theory of preferences*. Any re-interpretation of preferences in psychological or mental terms – such as the revival of hedonistic utility recently promoted by Kahneman and others (1997) – therefore must be based on considerations that are independent from those that motivated the development of Prospect Theory.

5. When the causal basis varies: intransitive preferences

My rejection of the mentalistic interpretation of preferences is based on a distinction between two different levels of explanation. And the distinction between the two levels is justified by multiple realizability. But multiple realizability is an empirical matter: what if the causal bases of preferences were *always* psychological, after all? 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. In this case, there does not seem to be any reason to deny that preferences are psychological dispositions.

Some contemporary attempts to identify ‘experienced utility’ or ‘utility in the brain’ are probably based on this sort of reasoning.²⁰ If not in the mind/brain, where could utility possibly be? The question seems rhetorical if we interpret choice theory narrowly as a theory of human decision making, and if we accept the internalist prejudice that pervades contemporary behavioural economics. But if these assumptions are challenged,²¹ it is easy to see that (i) there may be different causal bases for the preferences of different agents, and (ii) not all the bases are necessarily psychological.

I will proceed again by way of an example. Experimental and behavioural economists have demonstrated convincingly that preferences in some circumstances may be intransitive, 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 more concrete let us suppose that the choice concerns three types of car characterised by different degrees of comfort, speed, and price. Now, let us

19 E.g. Blackford et al. (2010)

20 See e.g. Kahneman et al. (1997), Glimcher et al. (2005), and – for a critical perspective – Vromen (2010), Fumagalli (2013).

21 For a forceful critique of both assumptions, see Ross (2005).

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 characteristic 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 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 the simple heuristic thus would generate intransitive preferences: X is preferred to Y, Y is preferred to Z, but Z is preferred to X.²²

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 justified.

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 very 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

²² 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).

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 well-known example of ‘Condorcet voting’, a problem that has been discussed extensively in choice theory.²³ Condorcet voting is an interesting case of intransitivity generated by voting rules. A non-psychological mechanism here produces an peculiar pattern of preferences. Although formally analogous, the mechanism thus is substantially different from the one that explains intransitivities 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 an institutional rule, rather than by a heuristic.²⁴ 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 features of the environment, for example the rules that force the committee to follow a majority-voting procedure.

In this kind of *externalist* explanation, the preferences of the committee are determined to a large extent by some feature of the environment in which the committee members have to make a decision. The causal basis of preferences is not always, or not even typically, constituted by psychological processes.

6. Taking stock

The controversy on the interpretation of utility 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

23 See e.g. Farquharson (1969), List and Pettit (2011).

24 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).

particularly evident during the last three decades, since the rise of behavioural economics has led to the development of psychologically-informed models in mainstream economics.

Seen in this light, the resistance against mentalism may seem 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 *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. In some circumstances, especially in the traditional domain of human decision-making, the causal basis is indeed mainly constituted by psychological mechanisms. But in other cases it is not. This explains why some economists resist the narrow identification of choice theory with the psychology of decision-making.

This 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 this field of science.

This brings an end to the main argument of this paper. With multiple realizability, the critique of the behavioural/mental dichotomy is complete. Multiple realizability however presupposes a ‘wide’ interpretation of choice theory – as a theory that is applicable to the behaviour of decision-makers that are not necessarily human. For those who do not find this presupposition persuasive, the rest of the paper is devoted to dispel a few worries and objections. In the next section I address the familiar complaint that a wide interpretation makes the theory of choice empty of explanatory content. In the final section, I examine the claim that behaving consistently with the principles of choice theory is a mark of the mental – and thus that any economic agent necessarily has a mind.

7. First coda: the domain of choice theory

There is no doubt that, historically, human decision-making has been the core domain of application of microeconomics. The modern theory of rational choice is a direct descendant of Neoclassical consumer’s theory, and its extension to non-human behaviour is a relatively recent development. But today choice theory can be and is routinely used to explain the behaviour of non-human animals, organizations, and even artefacts. One may worry that these extensions of the theory compromise its explanatory power. What is the purpose of attributing preferences and beliefs to robots, political parties, or fish? Political parties and robots may behave ‘as if’ they had preferences, but surely they do not really *want* anything in

particular. It is *us*, the observers, who find it useful to postulate such theoretical constructs; but the latter do not really refer to preferences in the genuine sense of the term.

The problem with this argument, of course, is that one needs to explain what ‘genuine’ means. If by ‘genuine’ we mean a state generated by mental or psychological mechanisms such as those that are implemented in human brains, then committees do not have preferences *in this sense*. But since the issue that we have started from is whether preferences (including those of human decision-makers) should be interpreted as mental states, clearly this objection would beg the question.

An alternative way to proceed is this: instead of searching for human-like mental states (whatever they may be) in non-human decision-makers, we should ask whether the attribution of preferences to any (human or non-human) agent provides useful information about its behavioural dispositions. Are preference-based explanations entirely *ad hoc*, or do they provide information that other kinds of explanation fail to deliver? If they were *ad hoc*, clearly the theory would be void of substantial content and of little scientific interest.

But choice-theoretic explanations are not empty. They provide useful information about the behaviour of creatures that are (1) consequence-driven, (2) motivated to pursue different goals, and (3) able to compare the value of such goals.²⁵ Another way to put it is that choice theory is valuably applied to creatures, artefacts, or organizations who possess a ‘control centre’ where information is processed, consequences are evaluated, and trade-offs about costs and benefits are made. How the control centre works, or what it is made of, is an issue that may influence *which* model of choice is likely to work best in each specific case, but that does not preclude the application of choice theory *per se*.²⁶

It may be helpful to review a few cases that fall outside the domain of the theory, beginning with behaviour that is not *consequence-driven* (condition 1). When the doctor taps my knee with a little hammer, my thigh muscle stretches and causes a jerk in the knee. Although the movement may be described in choice-theoretic terms, such a description would be *ad hoc* and would mischaracterize the process. The knee-jerk reaction is a reflex: it does not result from a process of evaluation of the consequences of my action.

Multiple goals (condition 2): reflex behaviour is not always easily distinguishable from other simple forms of decision-making. Russell’s famous chicken, who runs toward the feed until the day its neck is wrung, is able to evaluate only *one* consequence of its action. But empirically there does not seem to be much difference between mono-consequentialist choice and reflex behaviour: the chicken in a very obvious sense cannot help but run toward the farmer every time it sees him approaching, because it has been programmed (or has learned) to evaluate only this tiny aspect of the complex life of the farm.

25 There are clear analogies with Robbins’ (1932) famous definition of economics as ‘the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses’: consequence-driven behaviour is behaviour directed toward ends, and alternative uses are analogous to multiple goals. For a historical perspective on Robbins’ definition, see Backhouse and Medema (2009).

26 For a similar take on the importance of having a control centre (or ‘integrated workspace’, in his terminology) see Godfrey-Smith (2013).

Weighing (condition 3): a slightly more intelligent and diffident creature would take into account the probability that the farmer is going to string its neck. But this would require the capacity to compare different outcomes – such as being hungry and being dead – together with their relative probabilities. The choice-theoretic apparatus works best when we are dealing with a system that pursues different (and potentially conflicting) goals, guided by different information. The decision-maker must be able to resolve conflicts between her goals by weighing the pros and cons of different courses of action.²⁷

Most human behaviour does, of course, involve the evaluation of multi-attribute consequences. Humans pursue a variety of goals simultaneously, and often face decisions that require the making of trade-offs between different ‘goods’ and ‘bads’. A prominent human way to make such trade-offs is to think about the consequences of the actions and to engage in a systematic process of comparative evaluation. But conscious deliberation is by no means the only or the most important way to make an economic decision. Whether the control centre is in one’s brain, one’s computer, or in one’s stomach is of secondary importance.²⁸

That’s why the behaviour of many non-human systems can be explained in choice-theoretic terms. 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 (Appel and Elwood 2009). Even jellyfish, as we shall see shortly, have a control centre that “coordinates responding when stimuli are tending to elicit more than one response” (Albert 2001: 474).

8. Second coda: functionalism and mentalism

Endorsing a wide interpretation of the domain of choice theory makes the multiple realizability thesis plausible. At the same time, the wide interpretation does not make the theory empty or trivial. But, some philosophers may argue, a wide interpretation may *defeat* the anti-behaviourist, anti-mentalist position that I am trying to defend in this paper. To realize why it might be so, it is useful to start from a distinction between two fundamental conceptions of the mental:

It is now a commonplace, at least among philosophers, that there are two fundamentally different conceptions of mind and mental states. One is that the mind is distinguished by its intrinsic nature or character: by what it is in itself. It is different from matter because it is made out of different stuff, as

²⁷ Both the goals and the weighing processes must be stable enough to allow the prediction of behaviour.

²⁸ See e.g. Zajonc (1980). Those who fail to account for the multiple causal bases of preferences typically end up defending an over-intellectualised view of preferences. See for example Hausman (2012), and Angner (unpublished) for a critique.

with Descartes' *res cogitans*; or at least it has different intrinsic properties, such as being phenomenally conscious [...].

The other conception is that mind is distinguished, not by what it is intrinsically but by what it does extrinsically: by the roles it plays, the functions it discharges. On this conception, to have a mind is simply to function in a certain way. (List and Pettit 2011: 170-171)

The conception that may threaten the anti-behaviourist and anti-mentalistic position is the second one. According to functionalists,

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)²⁹

The functionalist approach is not just a philosophical curiosity. On the contrary, it often informs the way in which scientists think and talk about the mind. According to David Albert, a scientist working at the Roscoe Bay Marine Biology Laboratory in Vancouver, for example, some species of jellyfish display the typical characteristics that we normally associate with the capacity to make choices. The behavioural repertoire of Moon Jellies (*Aurelia* or *scyphozoa*)

include swimming up in response to somatosensory stimulation, swimming down in response to low salinity, diving in response to turbulence, avoiding rock walls, forming aggregations, and horizontal directional swimming. (...) The central nervous system of *Aurelia* sp. coordinates motor responses with sensory feedback, maintains a response long after the eliciting stimulus has disappeared, changes behaviour in response to sensory input from specialized receptors or from patterns of sensory input, organizes somatosensory input in a way that allows stimulus input from many parts of the body to elicit a similar response, and coordinates responding when stimuli are tending to elicit more than one response. (Albert 2011: 474)

Albert's article is titled 'What's on the mind of a jellyfish?'. It suggests that the mechanisms that govern the responses of *Aurelia* to different stimuli justify the attribution of mental states to jellyfish. But if

²⁹ Dietrich and List (2016) do not endorse functionalism, but a weaker position according to which playing a preference-role is merely 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.

any creature that exhibits choice behaviour in this sense deserves to be attributed a mind, then every preference *is* a mental state – contrary to what I have been arguing so far.

This argument is based on a profound equivocation of the sense in which ‘mentalism’ is relevant for economics. As mentioned earlier, the debate in economics concerns mainly the role that psychological data and psychological explanation should play in choice theory. It concerns a way of doing science – a scientific programme – rather than the philosophical problem of defining the mental. It is worth remembering that functionalism originates from a scientific programme too, rather than from the need to resolve a conceptual issue. The main goal of this programme – ‘Artificial Intelligence’ – was to understand cognition by modelling the mind as a computational system. 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.³⁰

Far from being merely terminological, the point of functionalism was to convince philosophers and scientists that it is possible to understand complex cognition without ‘looking under the hood’. More concretely, it promised to understand the mind even without an understanding of the way the brain works. The software can be studied independently of its hardware. Thus functionalism was motivated primarily by a *methodological*, rather than by an ontological concern.

Notice how different is the way in which mentalism is employed in the choice theory controversy. Mentalistic interpretations of preferences have been brought to the fore by the rise of behavioural economics. But 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 and neural basis of economic decision-making. If there is a methodological commitment, it is exactly the opposite than the functionalist’s commitment in AI and the philosophy of mind.

The main problem with a functionalist defence of mentalism then is that it would be disconnected with any specific scientific programme. It is not scientifically controversial that *something* determines the choices of an economic agent. The scientific controversy concerns what that ‘thing’ may be, and as we have seen it is likely that it differs from case to case. In some case it may be a neuro-psychological process, in another it may be something completely different – as when we are dealing with committees or jellyfish. Given this diversity, it seems unlikely that a single scientific theory will be able to account for the causal basis of preferences across the whole domain of application of choice theory. This is one of the reasons why economists prefer to be non-committal about the nature of preferences. A functionalistic mentalism may appeal to philosophers, but does not match scientific practice.

Another way to put it is to say that there is no ‘science of the mental’ in the functionalist sense. Or rather: the only science of the mental would be choice theory. But then to say that preferences are mental

30 See e.g. the essays on the mind-body problem collected in Putnam (1975).

would be akin to saying that preferences are preferences – it would not make any difference in any substantial, methodological sense. Behavioural economics in contrast is a scientific programme committed to study the properties and the causes of human preferences (among other things). It is committed to the claim that some preferences (those of human decision makers) have certain specific features in virtue of the underlying psychological processes that govern choice behaviour. This is a substantial hypothesis that makes a difference for the way in which economic science is and will be practised. Unfortunately, the explanations that behavioural economics puts forward are unlikely to be applicable to jellyfish or committees. In different domains different scientific theories are likely to be relevant. But this is something that a scientific account of preferences must be ready to cope with.

When economists argue about ‘mindless’ or ‘mindful’ economics, in contrast, something very important is at stake. The disagreement revolves around the use of psychology in economics. It has important implications regarding the training of young economists, the theories and the methods they should be taught, and which ones will help us make progress in the modelling and explanation of choice behaviour.

Economists do not argue about preferences as ‘mental state’ in the functionalist sense. This is something about which there is no disagreement, because it does not make any difference for the way in which economic science is done.

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. (unpublished) What are preferences?
- Elwood, R. W. & Appel, M. (2009) Pain experience in hermit crabs? *Animal Behaviour* 77: 1243-1246.
- Backhouse, R. E. and Medema, S. G. (2009) Defining economics: the long road to acceptance of the Robbins definition. *Economica* 76: 805-820.
- Blackford, J. U., Buckholtz, J. W., Avery, S. N., & Zald, D. H. (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, Mass.: Harvard University Press, 261-325.
- Bridgman, P. W. (1927) *The Logic of Modern Physics*. New York: Arno Press.
- Bruni, L. & Sugden, R. (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 & A. Schotter (eds.) *The Foundations of Positive and Normative Economics*, New York: Oxford University Press, pp. 43-69.
- Clarke, C. (2016) Preferences and positivist methodology in economics. *Philosophy of Science* 83: 192-212.
- Corbetta, M. & Shulman, G. L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* 3: 201-215.
- Cozic, M. (unpublished) Yet another attempt at clarifying revealed preference theory.
- Dietrich, F. & C. List (2016) Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics & Philosophy* 32: 249-281.
- Dowding, K. (2002) Revealed preference and external reference. *Rationality and Society* 14: 259-284.
- Farquharson, R. (1969) *Theory of Voting*. Oxford: Blackwell.
- Fumagalli, R. (2013) The futile search for true utility. *Economics and Philosophy* 29: 325-347.
- Glimcher, P. W., Dorris, M. C. & Bayer, H. M. (2005) Physiological utility theory and the neuroeconomics of choice. *Games & Economic Behavior* 52: 213-256.
- Godfrey-Smith, P. (2013) Cephaloids and the evolution of the mind. *Pacific Conservation Biology* 19: 4-9.
- 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.
- Gul, F. & Pesendorfer, W. (2008) The case for mindless economics. in A. Caplin & A. Schotter (eds.) *The Foundations of Positive and Normative Economics*, New York: 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 & A. Schotter (eds.) *The Foundations of Positive and Normative Economics*. New York: Oxford University Press, pp. 125-151.
- Hausman, D. M. (2012) *Preference, Value, Choice and Welfare*. New York: Cambridge University Press.
- Hicks, J. & Allen, R. G. D. (1934) A reconsideration of the theory of value. *Economica* 1: 52–76, 196–219.
- Holm, L., Ritz, C. & Ladewig, J. (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* 17:159–74.
- Hyman, G.M. (2009) *Addiction: A Disorder of Choice*. Princeton: Princeton University Press.

- Koob, G.F. and Le Moal, M. (2006) *Neurobiology of Addiction*. London: Elsevier.
- Kahneman, D. & Tversky, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263-291.
- Kahneman, D., Wakker, P. P. & Sarin, R. (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. & Pettit, P. (2011) *Group Agency*. New York: Oxford University Press.
- Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995) *Microeconomic Theory* (Vol. 1). New York: 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.
- Moscatti, I. (2013a) How cardinal utility entered economic analysis, 1909-1944. *European Journal of the History of Economic Thought* 20: 906-939.
- Moscatti, I. (2013b) Were Jevons, Menger and Walras really cardinalists? On the notion of measurement in utility theory, psychology, mathematics and other disciplines, 1870-1910. *History of Political Economy* 45: 373-414.
- Moscatti, I. (2015) Austrian debates on utility measurement, from Menger to Hayek. In R. Leeson (ed.) *Hayek: A Collaborative Biography, Part IV*. New York: Palgrave Macmillan, pp. 137-179.
- Mumford, S. (1998) *Dispositions*. Oxford: Oxford University Press.
- Nagatsu, M. and Poder, K. (unpublished) Economists' concept of choice: an experimental study.
- 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 Libreria.
- Putnam, H. (1975) *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Robbins, L. C. (1932) *An Essay on the Nature and Significance of Economic Science*. London: Macmillan.
- Rosenberg, A. (1993) *Economics: Mathematical Politics or Science of Diminishing Returns?* Chicago: University of Chicago Press.
- Ross, D. (2005) *Economic Theory and Cognitive Science*. Cambridge, Mass.: MIT 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: 41-27.

- 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: 355–83.
- Sen, A. K. (1977) Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy & 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., May, P., Reinikainen, K. & Näätänen, R. (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. & Kahneman, D. (1992) Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5: 297-323.
- Vanderbeeken, R. & Weber, E. (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. & McDougall, W. (1929) *The Battle of Behaviorism*. New York: Norton.
- Woodward, J. (2008) *Making Things Happen*. New York: Oxford University Press.
- Zajonc, R. B. (1980) Feeling and thinking: preferences need no inferences. *American Psychologist* 35: 151-175.