In Defence of Revealed Preference Theory

Johanna Thoma*

February 7, 2020

Abstract

This paper defends revealed preference theory against a pervasive line of criticism, according to which revealed preference methodology relies on appealing to some mental states, in particular an agent's beliefs, rendering the project incoherent or unmotivated. I argue that all that is established by these arguments is that revealed preference theorists must accept a limited mentalism in their account of the options an agent is modelled as choosing between. This is consistent both with an essentially behavioural interpretation of preference and with standard revealed preference methodology, and does not undermine the core motivations of revealed preference theory.

Keywords: Revealed Preference Theory; Preference; Expected Utility Theory; Mentalism; Behaviourism

*Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, United Kingdom. E-mail: j.m.thoma@lse.ac.uk URL: https://johannathoma.com/

1 Introduction

The theory of rational choice that is at the heart of much of orthodox economics is formulated in terms of preferences. Economic models in most of microeconomics, as well as much of macroeconomics typically describe agents as expected utility maximizers. Methods for the empirical measurement of demand functions, the rate of inflation, and other important economic variables presuppose that agents are expected utility maximizers. What this has standardly been taken to mean, since about the middle of the 20th century, is that agents have preferences which fulfil various conditions, making it the case that they can be represented as expected utility maximizing.

This paper is concerned with the cluster of views commonly referred to as 'revealed preference theory'. At the core of revealed preference theory as I will understand and defend it in the following is a claim about the interpretation of 'preference' in standard economic theory – both when it is applied in economic modelling, as well as, and especially, in empirical estimation and measurement. Revealed preference theorists are committed to a behavioural understanding of preference, whereby they equate preference with actual or hypothetical choice behaviour.[1] According to that understanding, when an economist ascribes a preference for an option a over an option b to an agent (be it a real or a model agent), this is meant to capture nothing more than that the agent does or would choose a over b from some specified choice set – for instance, that she does or would choose a over b when only those two options are available, or that she does or would not choose b in any choice situation where both are available.[2] The behavioural interpretation of preference gives a technical meaning to preferences as they appear in standard economic theory, which consciously departs from the everyday understanding of preference as conative mental attitude.[3]

Apart from a general commitment to the behavioural understanding of preference in economic theory, revealed preference theory is characterized by a methodological programme that aims to (i) use past choice data to ascribe preferences and utilities to agents, (ii) use these preference and utility ascriptions to make predictions about unobserved choices, and (iii) lay the formal foundations for the mapping of choice behaviour to pref-

[1] This view is very widespread in economics. See, for instance, Savage (1972), Luce and Raiffa (1957), Harsanyi (1977), Mas-Colell et al. (1995), Binmore (2008), Gul and Pesendorfer (2008) and Gilboa (2009). For a defence in the philosophical literature, see Maher (1993).

[2] I will not take a stance here on which of these behavioural interpretations should be adopted. A potential problem with the first is that it is not clear whether it allows us to express indifference. A potential problem with the second is that it conceptually rules out various violations of standard expected utility theory.

[3] Note that according to this understanding of the behavioural interpretation of preference, if what an agent actually or hypothetically chooses when confronted with a particular choice set is not stable over time, then the agent does not have stable preferences – it is not the case that she fails to have preferences at all, or that she chooses counter-preferentially.

erence. Under a behavioural interpretation of preference, no inference about conative mental states (preference in the ordinary sense) is made when going from observation of choice behaviour to preference and back to a prediction about choice. On the face of it, we are merely making predictions about future choice on the basis of past choice, under the assumption that various consistency conditions hold, with preference and utility assignments serving as a useful shorthand. Revealed preference theorists have thus taken this methodology to be more inductively safe, to find more general application than methods that make more explicit appeal to mental states, and to allow greater separation from the psychological sciences. These alleged advantages depend on the behavioural interpretation of preference, and the behavioural interpretation is in turn bolstered by success of the methodology (if, indeed, it is successful).

Insofar as it is committed to a behavioural understanding of preference, revealed preference theory is widely rejected in the philosophical literature.[4] It is frequently associated with long discredited forms of behaviourism and positivism. Recently, Clarke (2016) has shown persuasively that the behavioural interpretation of preference as it features in economic models, or what he calls the 'shorthand view' does not imply any problematic behaviourist or positivist thesis. And he and others have provided compelling arguments to favour a behavioural interpretation, or at least to move away from the interpretation of preference as mental state usually proposed as an alternative.[5] But one line of criticism of revealed preference theory, brought forward most forcefully by Hausman (2000), has proven especially sticky. According to this line of criticism, the empirical success of revealed preference methodology in fact ultimately relies on appealing to some mental states, in particular the agent's beliefs, and her own conception of the options open to her. This is then either taken to show that revealed preference theory is incoherent, or at the very least that it is unmotivated. Alternative interpretations of preference, usually as a kind of mental state, are proposed instead. Even Guala (2019), who rejects the interpretation of preference as mental state, and whose argument I am largely sympathetic with, is persuaded by this line of argument to reject the revealed preference theorist's identification of preference and choice, and interpret preferences as belief-dependent dispositions instead.

Against this line of argument, this paper shows that a charitable interpretation of revealed preference theory, which preserves the identification of choice and preference, is possible, and faithful to economic practice.[6] Despite the popularity of the argument, a close analysis of exactly how revealed preference theorists must appeal to beliefs, and where this leaves the identification of preference and choice and the revealed preference theorist's methodological programme is in fact still wanting. And this is what I aim to

[4]See, in particular, Rosenberg (1992), Hausman (2000, 2012), Craver and Alexandrova (2008), Dietrich and List (2016), Bradley (2017), and Guala (2019).

[5]See, e.g., Angner (ming) and Guala (2019). I provide a further argument in Thoma (2020).

[6]Note that my argument will be limited to what is commonly referred to as 'positive' economics, and bracket the question of welfare and its relation to preference.

provide here. I will conclude that revealed preference theorists need to concede only a minimal kind of mentalism, which leaves the identification of preference and choice, and thus, I will argue, the behavioural interpretation of preference, untouched. The root of the critics' concern is that revealed preference theory is only empirically successful if we can attribute to the agents we wish to study preference relations that are generally consistent in the way the theory requires, and generally stable over time. This paper will concede that revealed preference theory can only hope to be empirically successful in this way if agents' options – the objects of both preference and choice – are described in a way that is consistent with how they present themselves to them. Getting the description of options right will thus often involve making assumptions about agents' beliefs. However, I will argue that this only amounts to mentalism in our theory of options, and a limited kind of mentalism at that. Such mentalism is consistent both with the behavioural interpretation of preference and with standard revealed preference methodology. In fact, I will provide evidence that current economic practice is by and large consistent with the concession of a limited kind of mentalism about options.

Moreover, I argue that combining a behavioural interpretation of preference with a limited mentalism about options does not undermine the core motivations its proponents have cited in favour of revealed preference theory: Revealed preference theory still largely black-boxes the psychological processes that lead to choice, and thereby achieves greater generality, avoids controversial substantive commitments about psychological processes we know little about, and preserves clearer disciplinary boundaries than expected utility theory would under more mentalistic interpretations of preference. In addition, revealed preference theory still eliminates several problematic inductive steps in the measurement of economic variables and prediction of economic choice behaviour, in particular the inductive steps from choice to mentalistic preference and vice versa.

Revealed preference theory should thus not be as quickly dismissed as it often is in the philosophical literature. Its core methodological commitment to the identification of preference and choice, the motivations commonly cited in favour of it, and economic practice are coherent with each other, and not undermined by the observation that mental states must have some role to play in revealed preference theory. There may, of course, be other reasons to reject revealed preference theory, and a fuller defence of the view must be provided elsewhere.[7] However, I believe my argument removes the need for those who, like Guala, reject mentalism, to distance themselves from revealed preference theory. And it shows one of the core arguments often cited in favour of mentalism to be unsound.

[7]The arguments against mentalism cited in footnote 5 above, I believe, make a strong case in combination with my argument here.

2 Revealed Preference Theory

Before evaluating the adequacy of revealed preference theory, it is important to get a sense of how preferences feature in economic methodology. First, preferences feature heavily in models of individual choice. As mentioned in the introduction, the standard economic model of individual choice describes agents as expected utility maximizers. Various representation theorems provide 'axiomatizations' of expected utility theory, showing that an agent's preferences can be represented as expected utility maximizing in the framework proposed if and only if they abide by a number of axioms. The representation theorem most commonly appealed to in economics goes back to von Neumann and Morgenstern (1944). Most importantly, von Neumann and Morgenstern require agents to have complete and transitive preferences over outcomes and lotteries (probability distributions over outcomes), and for these preferences to abide by the independence axiom. Von Neumann and Morgenstern's representation theorem assumes probabilities over outcomes to be given (hence this is often called 'objective' expected utility theory), and shows that if and only if an agent's preferences abide by the axioms will there be some utility function over outcomes such that the agent prefers one lottery over another if and only if the expectation of utility associated with that lottery is higher. In contexts where probabilities cannot plausibly be assumed to be given, Savage's (1954) representation theorem for 'subjective' expected utility theory is usually invoked, since it allows us to also assign probabilities to the agent.

Since the rise to prominence of these representation theorems, economists have typically viewed utility as nothing more than a convenient device for representing an agent's preferences – cutting ties with the more substantive roots of the concept as a psychologically real quantity, such as happiness, going back to 19th century British moral philosophy. Revealed preference theory goes one step further in cutting ties with these mentalistic origins of utility theory by interpreting preference itself not as a mental attitude distinct from the choice patterns to which it gives rise, but rather equating it with the choice patterns themselves.

The possibility of an understanding of preference that reduces to choice is established by a further group of representation theorems, which show that if and only if an agent's choices over sets of options (including non-binary ones) abide by various consistency conditions can they be represented as optimizing according to some binary relation R which is complete and transitive over all options. Building on work by Samuelson (1938), this was first proven by Houthakker (1950), who introduced what is now known as the 'Strong Axiom of Revealed Preference'.[8] Under uncertainty, additional axioms guarantee consistency

[8]The revealed preference theorem most commonly appealed to today is due to Afriat (1967), which employs the 'Generalized Axiom of Revealed Preference'.

with expected utility maximization, within the objective and subjective framework.[9] The revealed preference theorist goes on to call R the preference relation: Preference is whatever binary relation conveniently represents the agent's choices in the way captured by the theorems – just as utility is standardly interpreted to be just whatever conveniently represents preferences. As is generally acknowledged, the theorems do not prove that choices reveal preferences in the ordinary sense of preference, as conative attitude. They merely show that we can define a technical notion of preference that reduces to choice, which may or may not track ordinary preference. Revealed preference theorists claim that when expected utility theory is applied in positive economics, preference should be understood in the technical way.[10]

While these representation theorems are an exercise in pure theory, they form the basis for empirical methods that allow for the estimation of individual demand functions (relating the quantities purchased of some good to its price) and Engel curves (relating the quantities purchased of some good to income level) from a limited set of observations of purchasing decisions at different price points and income levels.[11] Economists can use these methods to describe and explain the choices of real individuals they have observed as expected utility maximizing, and to make predictions of future and yet unobserved choices. In addition to the assumption that an agent's actual and hypothetical choices at any one point in time obey by the axioms of the representation theorems, these methods rely on the assumption that an agent's choice behaviours remain stable over time.[12]

Revealed preference theory, as I understand it here, is committed to the empirical

[9]See Green and Srivastava (1986) for a revealed preference theorem in the objective framework, and Chambers and Echenique (2016), Chapter 8 for an overview of revealed preference theory under uncertainty. Though these find less application in practice, there are also revealed preference theorems for game theory, establishing the representability of joint choice functions by standard solution concepts. See Chambers and Echenique (2016), Chapter 10.

[10]Bernheim and Rangel (2008: 158), provide an especially clear statement of this position: "Though we often speak as if choices are derived from preferences, the opposite is actually the case. Standard economics makes no assumptions about how choices are actually made; preferences are merely constructs that summarize choices. Accordingly, meaningful assumptions pertain to choices, not to preferences. Though the terminology suggests a model of decision making in which preferences drive choices, it is important to remember that the standard framework does not embrace that suggestion; instead, R is simply a summary of what the individual chooses in a wide range of situations." (as also cited in Hands 2013)

[11]See, in particular, Afriat (1973) and Varian (1982).

[12]This assumption is sometimes treated as part of the behavioural interpretation of preference. See, for instance, Bermudez (2009). I want to resist this, since it would imply that agents whose choice behaviour is not stable over time fail to have preferences, making all of expected utility theory inapplicable to them. However, this would be throwing the baby out with the bath water. Such agents' actual and hypothetical choices can still abide by the axioms at any one point in time, and indeed they can exhibit stable choice behaviour for the most part, over long stretches of time. Treating stability of choice behaviour not as part of the definition of preference, but as an additional assumption of the revealed preference approach implies that instabilities of choice behaviour make prediction of agents' choices with revealed preference methodology fallible, rather than in principle inapplicable.

methodology just described for (i) imputing preferences and utility and demand functions from choice data, and (ii) using these to make predictions about unobserved choice behaviour and economic phenomena, based on the (iii) formal foundations laid by the representation theorems for mapping choices to preferences. Commitment to this methodology is coupled with the conceptual claim that preferences are to be understood behaviourally, as mere convenient descriptions of an agent's choice behaviour. Some central motivations for advocating this package are canvassed in Section 6 below. Note that while revealed preference theorists push for the centrality of their methodology in empirical economics, they of course acknowledge that there is more to economics than this methodology. And they typically take their conceptual claim, the behavioural interpretation of preference, to extend beyond parts of economics that use this empirical methodology. Indeed, we can also extend it to more theoretical parts of economics.

Models in theoretical economics typically posit some expected utility maximizing agents, often with simple utility functions sensitive only to the agent's own profits or some combination of the agent's own wealth, consumption or leisure. Unlike in more empirical parts of economics, the assumption of expected utility maximization is not applied to real agents in the economy, but to theoretical entities that play a certain role in models, and that do not have a clear and definite analogue in the real economy. Akerlof's (1970) famous model of a market for used cars, for instance, features two types of traders of used cars, both of which are portrayed as expected utility maximizers whose utility functions are sensitive only to wealth levels and some measure of the quality of any cars they own. These groups of traders are not plausibly understood to be direct representations of any real used car traders.

Exactly how and what we are supposed to learn from such theoretical models about the economy is a matter of dispute within philosophy of economics. I agree here with, e.g., Sugden (2000, 2009) and Gilboa et al. (2014) at least in what these models are trying to offer: Illustrations of a general type of mechanism that is instantiated in many different ways in different economic settings. Insofar as there are relevant similarities between the model and some economic setting or phenomenon, we are licensed to draw some inferences from the model. This is supposed to be possible even if the agents featuring in the model are not direct analogues of any particular real world agents. The conceptual claim of revealed preference theory, that preferences are to be understood as actual or hypothetical choices, applies to such theoretical models, too. According to this claim, positing expected utility maximizing agents in theoretical economic models amounts to positing theoretical entities that choose in a certain way. It does not amount to positing theoretical entities whose choices are caused by some internal state in a specific way. And so, whatever else we might infer from them, theoretical economic models featuring expected utility maximizing agents never license inferences about the mental causes of the choices of real agents.

3 The Anti-Behaviourist Challenge to Revealed Preference Theory

Revealed preference theory is often described as an outgrowth of an outdated and discredited positivist philosophy of science, characterized by a general unwillingness to posit unobservable entities, in particular mental states.[13] It is now generally acknowledged that no science can make progress without positing unobservable entities. Psychology, in particular, has largely abandoned behaviourism, the view that banished appeal to unobservable mental states in favour of analysis in terms of patterns of observable behaviour. Since positivism and behaviourism are untenable, it is argued, revealed preference theory must also be.

Clarke (2016) shows convincingly that the identification of preference and patterns of choice can be isolated from, and can be motivated independently of discredited positivist and behaviourist theses. Identifying preference and choice does not, for instance, thereby commit one to the idea that the aim of science in general is to explain the observable in terms of the observable, that choice can be observed in a theory-independent way, or to the idea that choice data is the only kind of data admissible in economics. As will become clear in the following, empirical revealed preference methodology also does not commit economists to these positivist theses. However, there is one particularly sticky issue not addressed by Clarke, that has kept even those otherwise sympathetic, such as Guala (2019), from endorsing a behavioural interpretation of preference. And that is the idea that preference attribution is always inextricably linked to the attribution of mental states, in particular beliefs, and that this makes revealed preference theory either incoherent or unmotivated.[14]

Consider the following example. Suppose an economist and her friend visit a sushi restaurant for the first time. The economist has read about wasabi being very spicy and knows what it looks like. Her friend mistakes it for avocado and devours a whole spoonful. If the economist models her friend's choice options as "eating a spoonful of wasabi" and "not doing that", then as a revealed preference theorist, she will conclude that her friend prefers "eating a spoonful of wasabi" to "not doing that". As a revealed preference theorist, she may predict her friend will choose in the same way on future occasions. After all, as we have seen above, expected utility theory is committed to the claim that agents have preferences that are consistent in the way described by the standard axioms, and application of empirical revealed preference methodology used to impute demand functions requires an assumption that agents' preferences are stable over

[13] See, for instance, Rosenberg (1992), Hausman (2000, 2012), Craver and Alexandrova (2008), Dietrich and List (2016), Bradley (2017).

[14] See, in particular Hausman (2000, 2012) on this.

time. However, unless her friend has very unusual tastes, the economist will inevitably find that her prediction turns out false, and the revealed preference approach will have failed her.

The problem here clearly is that we have somehow not taken proper account of the fact that the economist's friend initially has a false belief about the option in front of her. Critics have concluded from this kind of case that preference should not be identified with choice: The agent chooses wasabi (due to a false belief), but does not prefer it. However, I take it that a natural reaction by an economist to this case would be that the decision problem has been misspecified, already at the level of the description of choice. That is, the agent should not be described as choosing wasabi. For one, misspecification of the objects of choice is the right diagnosis in a related class of cases. Take, for instance, the following variations of our case: Suppose the economist describes her friend's options in very general terms as simply "eating food" and "not eating food". Or suppose she is specific only in ways that capture only the superficial qualities of the options, for instance by describing her friend as choosing between "eating a thick green paste" and "not eating it". In both cases, she would again be unlikely to uncover a consistent and stable preference relation when she further observes her friend: She may observe her friend choosing to eat other food just moments later, and she may observe her eating a thick green paste with her tortilla chips just the next day.

In these two cases, I think it is uncontroversial that the objects of choice have been misspecified. If revealed preference methodology is to be empirically adequate, which requires the identification of a stable and consistent preference relation, then it must abide by some standard for the specification of an agent's options that will yield this verdict. I want to argue that revealed preference theorists should deal with the first case, involving false beliefs, in just the same way: By providing standards for the specification of choice options that help economists avoid the problem. As we will see in the next section, these will allow economists to retain the identification of preference and choice.

Even if critics were willing to go along with identifying the problem at the level of description of the objects of choice already, they may go on to argue that any adequate standard of option specification will have to make some reference to agents' mental states. And so a general kind of behaviourism, which disallows any reference to mental states, or even behaviourism constrained to revealed preference methodology, is untenable for revealed preference theorists. Let me sketch this argument applied to the cases just described.

To start with the first case, as just mentioned, the economist's mistake in describing her friend as choosing whether or not to eat wasabi would be that she did not correctly take into account what her friend does or does not know about her options. Her friend does

not know that the paste in front of her is wasabi. She might not even know what wasabi is, and that it is very spicy. It seems like, to avoid the above mistake, the economist must acknowledge in the description of options that her friend mistakenly thought the wasabi was something more delectable. Otherwise she will not be able to uncover a consistent and stable relation when observing her friend's future choice behaviour. Indeed, if the economist knew about her friend's mistaken beliefs, applying revealed preference methodology in the way described in the first scenario would be silly. Of course, the economist may not know about her friend's mistaken beliefs, in which case the misspecified decision problem may be her best attempt at uncovering a stable and consistent preference relation. But all this shows is that the revealed preference methodology is fallible whenever knowledge of agents' beliefs is imperfect. In any case, the success of revealed preference methodology relies on getting the agent's beliefs right.

In the second and third case, the economist does come up with a description of the options that is at least consistent with her friend's beliefs about her options. After all, her friend very likely does believe the wasabi to be food, and she very likely does believe it to be a thick green paste. Here, the problem seems to be that the economist has not described the options in a way that captures everything relevant to her friend's choice. Again, we might think that, in order to capture everything that is relevant in the appropriate sense, we have to consider the agent's mental states. We might, for instance, think that the economist should include in the description of the options all the features of the options that the agent finds desirable, such as the thick green paste's taste.[15] Or we might think that the economist should try to describe options in the way and at the level of detail that the agent herself conceives of them.[16] Either way, the economist appeals to further mental states, such as desires and mental representations of options.

If this analysis is correct, a general or even local kind of behaviourism is untenable for revealed preference theorists. There are two kinds of conclusions critics of revealed preference have drawn from this. First, if we define revealed preference theory in such a way that it is committed to strict behaviourism about revealed preference methodology, revealed preference theory appears to turn out incoherent: It must endorse a mentalistic theory of option specification while being committed to a behaviourism that should include one's theory of options.[17] Second, even if revealed preference theory is not taken to be committed to strict behaviourism by definition, revealed preference theory allegedly loses its core motivation if it must concede appeal to some mental states.[18]

[15]Pettit (1991: 165), for instance, argues that two outcomes (or, in the standard economic case, consumption bundles) should be distinguished from each other just in case they differ in terms of some property that is desired or undesired by the agent.

[16]We find this proposal, for instance, in Bradley (2017: 60).

[17]The argument in Hausman (2012: 28), for instance, relies on taking revealed preference theory to be committed to the view that preferences can be inferred from choices "regardless of belief".

[18]See, for instance, Bradley (2017: 60), Dietrich and List (2016).

The next sections aim to establish that it is uncharitable to define revealed preference theory in such a way that it is committed to behaviourism of a kind that forbids any mentalism in one's theory of options. Examples such as the ones just discussed do show that revealed preference theorists must adopt a kind of mentalism in their theory of what the options are that agents are choosing between – though the mentalism about options that must be conceded is more limited than critics claim. But this does not undermine the core conceptual commitment of revealed preference theory, namely the behavioural interpretation of *preference*. In fact, I will argue that the concession of a limited kind of mentalism about options is by and large consistent with economic practice, and does not render revealed preference theory unmotivated.

4 Mentalism about Options

The representation theorems for expected utility theory, and generally any application of the theory, start with some model of a decision situation. Von Neumann and Morgenstern's representation theorem assumes that agents have preferences over outcomes and/or lotteries. Savage has agents have preferences over assignments of outcomes to states of the world. The revealed preference theorems do not start with ascriptions of preferences, but with ascriptions of choices. Depending on the framework, these choices again have as their object outcomes, lotteries, or assignments of outcomes to states of the world. Before we can apply economic theory to a real world choice situation, we thus have to describe the choice situation in those terms. Here, there is room for mistake, and economists need some standards for the specification of the objects of preference and choice. My proposal is that revealed preference theorists should apply standards that help them in their core methodological programme, that of describing agents' choice behaviour with consistent preference relations, and using those to make predictions of future choice behaviour.

I can find no evidence that economists apply different standards for the specification of the objects of choice and the objects of preference – which would rule out the identification of preference and choice. For one, they use the terms 'options', 'outcomes', 'lotteries' etc. in both cases. Indeed, there seems to be no reason for revealed preference theorists to concede that the objects of choice and the objects of preference should be described differently if there are workable standards for the description of choice options that address the problems discussed in the last section, allow them to retain the identification of preference and choice, and more generally work for their core empirical project. I here present standards that I think fit the bill, and that are, moreover, consistent with economic practice.

For there to be a hope of uncovering a stable and consistent preference relation in the

way the revealed preference approach aims to do, the standards for the specification of agents' choice options must help revealed preference theorists avoid the kinds of mistakes discussed in the last section, at least in cases where economists are in the right epistemic situation to avoid them. I want to suggest the following standards for the specification of an agent's choice options in the ideal circumstances where we have full knowledge of the agent's hypothetical choice behaviours and beliefs:

1. The description of options should be consistent with the agent's beliefs about the nature and consequences of the actions open to her, provided the agent's relevant beliefs are mutually consistent.[19]

2. A perceived feature of a choice situation should be included in the description of the agent's options whenever that feature affects the agent's choice behaviour, that is, when there are choice situations where the agent would make a different choice when she believes that feature is present or absent respectively.[20]

Describing options in this way does not guarantee that standard revealed preference theory will be able to capture every agent's choice behaviour and make correct predictions using the standard methodology. Even if we describe options in this way, agents may still fail to display stability in their choice behaviour, for instance if their fundamental tastes change. Insofar as revealed preference theory relies on a stability assumption it thus remains fallible. Moreover, even if options are described in this way, agents may still fail to abide by all the standard axioms at any one point in time. For instance, they may still have cyclical preferences.[21] And so the above rule for specification under ideal circumstances does not make revealed preference theory infallible. However, it does rule out all the problematic cases discussed in the last section.

The first rule is designed to avoid the problem of false beliefs as in the first wasabi case — here the economist's mistake is that her specification of the options is not consistent with her friend's beliefs about them. The economist should either describe the options in a way

[19]'Belief' here should be understood in a permissive sense, as ascribable to even unsophisticated non-human or artificial agents that register things about their environments. I take no stance here on the special case where the agent's relevant beliefs are mutually inconsistent.

[20]Various authors have proposed that rules for the specification of options should be preference-based, in the sense that two options should be distinguished just in case the agent's preferences distinguish them (e.g. by the options being ranked differently against further options, or the agent having strict preferences between them). See, e.g., Joyce (1999: 52) or Dreier (1996). In effect, what I am proposing here is such a rule applied to a behavioural interpretation of preference.

[21]Indeed, this paper does not mean to offer a response to standard counter-examples to expected utility theory. However, insofar as alternatives to expected utility theory, built on less restrictive axioms (such as rank-dependent utility theories), have a better empirical fit with observed choice data, this paper can be read as providing arguments in favour of a behavioural interpretation of those theories, and the empirical methodology for predicting choice behaviour based on those theories. See, for instance, Harrison and Ross (2018) for an application of revealed preference methodology in a rank-dependent utility framework.

that is consistent with those beliefs, e.g. as "eating a delicious bite of avocado", or refrain from inferring a preference at all. One way an economist might achieve consistency of the option specification with the agent's beliefs is in offering either a very general, or a very superficial description of the agent's options – for instance, by describing the economist's friend's wasabi-eating options as in the second and third cases above, that is, as "eating food" or "eating a thick green paste" respectively. However, doing so would violate our second rule, as we would be leaving out many factors that presumably make a difference to the agent's choice, such as the food's taste.

The two rules also allow us to respond to a related worry that leads Guala (2019: 3-4), to reject the identification of preference and choice. He considers two agents each deciding between the same two pizza restaurants. They each 'prefer' a different one, but they go to the same restaurant, because one of them has a false belief about his 'preferred' restaurant being closed. Guala claims that in this case the two agents' choices are the same, but we should not conclude that their preferences are also the same. And so we cannot identify preference and choice. Applying the theory of options presented here, our reaction should be that their choices are in fact not the same: There is no way of describing the options identically in both cases, while preserving consistency with each agent's beliefs and capturing everything relevant to their choice.

Economists are of course never in the ideal situation where they have full knowledge of the agent's hypothetical choice behaviours and beliefs. And in fact, if they were, revealed preference methodology would be of much less use. Given economists' actual epistemic limitations, the above standards can only be approximated. In practice, what economists should thus try to achieve is firstly, consistency with our best estimate of the agent's most relevant beliefs, that is, beliefs about features of the choice situation that we have reason to believe affect her choice the most. In the wasabi case, if the economist has good reason to believe that her friend cannot identify wasabi, and that, plausibly, this is highly relevant to her choice, then she should not describe her friend's option as "eating a spoonful of wasabi". The next section will look in greater detail at how this standard is applied in practice, but note here that consulting non-choice evidence can be highly relevant in applying this standard, especially when economists can't make a good guess about agents' mental states.[22]

Secondly, given their limited knowledge of agents' beliefs and choice behaviours,

[22]Note also that relevant false beliefs may be hard to spot even in the lab, as they may fail to show up as choice reversal in the lab when a subject learns about her mistake, for instance out of embarrassment. Part of the difficulty of applying this standard is distinguishing such a case – where failure to spot the false belief will lead to false prediction of future choice behaviour, as the belief is choice relevant outside of the lab – from a non-problematic case of stable choice not affected by false beliefs. This difficulty translates into uncertainty whether what we find in the lab has external validity. Such doubts about external validity are of course not unique to revealed preference methodology. The two rules described here acknowledge that non-choice data, in particular about beliefs, may be relevant for dealing with such difficult cases.

economists should strive to include a feature in the description of a choice problem whenever they have reason to believe that the agent's belief in the presence or absence of that feature significantly affects the agent's choice behaviour in the kinds of contexts the economists are interested in. Again, under normal circumstances, this will exclude describing the wasabi-eating option as "eating food" or "eating a thick green paste", as this excludes features of the options we know to affect choice. The last qualification to contexts of interest ensures that economists do not have to worry, for instance, about agents' choice behaviours being radically different in radically different economic systems, unless, of course, they are interested in studying the effects of the introduction of such an alternative system. And so they can omit a full description of the current economic background conditions in most applications.

I take the standards for the specification of choice options described here to adequately deal with the problematic cases introduced in the last section. They help revealed preference theorists represent choice behaviours with stable and consistent preference relations, retaining the identification of preference and choice. Moreover, I will argue in the next section that economic practice is already in line with these standards. They are therefore not revisionary, but offered as a sympathetic reconstruction of actual practice. These standards do concede that some appeal to mental states must be made in order to get the specification of choice options right. In particular, economists must make sure that the description of options is consistent with the agent's relevant beliefs. However, note that the standards I have proposed here are less demanding, in terms of the knowledge of or assumptions about mental states they require, than those suggested by critics of revealed preference theory. Consistency with the agent's relevant beliefs is all that is required in terms of direct appeal to mental states. We do not need to require that options are described in the way in which agents conceive of them, or in a way that captures all of their relevant desires. We only need to make sure we capture everything that significantly affects choice behaviour.

My proposal is thus that revealed preference theorists can continue to think of preference as a mere convenient representation of choice, as long as they model agents as choosing between options described in the partly mentalist way I have described. The crucial question now is whether the partly mentalist theory of options I offered is consistent with an interpretation of preference that is still recognizably behavioural.[23] I believe it is. Most importantly, with this theory of option description in hand, we can still think of preference as choice, as going for, or picking one option over another – just as choice between options described in a way that must take account of the agent's relevant beliefs. This is fundamentally different from preference as a conative attitude of one option over

[23]It may, of course, at the same time be consistent with a mentalist interpretation of preference – see in particular my discussion of functionalism below. But my main purpose is to show that a charitable and essentially behavioural interpretation of revealed preference theory is possible.

another, as on the ordinary understanding of preference.

One might think now that we have simply exploited an ambiguity in the notion of choice, which can be given a more behavioural reading (as bodily movements that bring about some outcome), and a more mentalist one (e.g. as intentional choice) in its own right.[24] Perhaps we have just insisted on a thoroughly mentalist notion of choice, thereby preserving the identification of preference and choice, but sacrificing what was distinctively behavioural about revealed preference theory. Granting revealed preference theorists the labels 'behavioural interpretation of preference' and reference to 'choice behaviour' would then be misleading.

But the theory of options proposed here does not turn the notion of choice in revealed preference theory into intentional choice, where an agent's intentions are captured by the description of the options she is choosing: The theory of options we have proposed does not aim at capturing options as the agent herself conceives of them, or intends them to come about. It merely requires consistency of the option description with what the agent believes. While the option description we end up with may sometimes match the content of an agent's intentions, this needn't be the case, and revealed preference theory, as I sketched it, also allows us to capture unintentional choices – as long as these unintentional choices are consistent with the EUT axioms.

As mere behaviour does not have an object in any straightforward way, but choice, as it features in economists' formal frameworks, does, economists were always in need of some theory of option description when capturing behaviour in terms of choice. When that theory of option description is ultimately geared towards a framework that helps us to accurately capture and predict behaviour and nothing more, and is non-committal about the mental causes of behaviour, I believe the resulting notion of choice deserves to be called 'behavioural'. And what aiming at fulfilling the two conditions specified above allows revealed preference theorists to do is capture the behaviour of agents as consistently responding to what they perceive or believe about their choice environments (where their behaviour allows of such an analysis), which has proven very useful in predicting the behaviours of many different kinds of agents, non-human or human.

Some functionalists, such as Dietrich and List (2016) may hold that we have nevertheless characterized a concept of preference that plays the functional role of desire or preference in the ordinary sense: That which combines with belief in order to explain choice behaviour. In that case, the account I have offered would be consistent with a functionalist kind of mentalism about preference that ascribes mental states in virtue of this function. On my account, preference obviously does not explain *choice* in a folk psychological way, given we have identified preference with patterns in choice, and choice

[24] See Clarke (2016: 201) on this distinction.

is itself belief-dependent. But one might think it explains behaviour, when we think of behaviour as the mere description of outward bodily movements.

I argue in detail elsewhere ([redacted]) that this kind of functionalism about preference in decision theory (not functionalism in general) is misguided, because preference as we have characterized it here does not combine with belief to explain choice behaviour in the way desire and belief do in ordinary folk psychological explanations. So let me just point out here that insofar as my aim is to show that a charitable (that is, neither incoherent nor unmotivated) and still distinctly behavioural interpretation of revealed preference theory is possible, the possibility of this kind of functionalism does not undermine my argument. Revealed preference theory under my interpretation may be attractive even to those who are not functionalists of this kind, and will thus not interpret preference as we have characterized it mentalistically. Indeed, as we will see in section 6, revealed preference theory is usually motivated explicitly in opposition to mentalism about preference, aiming to black-box the mental causes of choice. Moreover, as we will see in the next section, while economic practice is consistent with limited mentalism about options, it still avoids ascription of mentalistic preference.

Guala (2019), rightly, I think, resists the functionalist interpretation of preference as a conative mental state. However, he argues that preferences should be understood as belief-dependent dispositions that play a characteristic functional role that makes them distinctly non-behavioural. However, his reasons for insisting that preference should be understood as non-behavioural do not apply to our account. One core reason, which we have just seen illustrated in the pizza example, is that he takes only preference and not choice to be mediated by beliefs.[25] On my account, both preference and choice are belief-dependent, but minimally enough to deserve the label 'behavioural'. Moreover, unlike folk psychological explanation, the main way in which Guala takes preferences to be explanatory does not rely on us denying that preferences are essentially behavioural. In particular, he thinks that preferences allow generalizations when the cause of choice is multiply realizable. Behavioural generalizations can also serve such a function. The following will show that preserving the identification of preference and choice in the way I have argued for here is the more charitable reconstruction of economic practice, and removes the need to distance ourselves from the revealed preference project.

[25] In fact, the only straightforward way of maintaining the distinction between choice and preference in his example would be to suppose that there are different standards for the description of choice options and the objects of preference. But that does not seem to fit economic practice.

5 Mentalism about Options in Practice

On the account offered in the last section, at least one fallible inductive step is involved when economists infer a preference from observing behaviour. Economists must come up with a characterization of the agent's options that is consistent with her relevant beliefs about them. If economists fail to do so, they may wrongly ascribe preference. Moreover, this fallible inference involves the economist making a judgement about the agent's mental states – about her beliefs about her options. However, while I think economists thus cannot escape a partly mentalistic theory of options, I argued that this is compatible with an essentially behavioural interpretation of preference. Below, I will argue that conceding a theory of options that is mentalistic in the limited sense described here preserves the main advantages economists have seen in revealed preference theory. For now, let me highlight that this concession is by and large consistent with economic practice.

I will start by looking at revealed preference methodology in the narrow sense, which uses choice data to attribute preferences to agents and then make predictions about unobserved choice behaviour. When attributing preferences to agents on the basis of choice data, beliefs play the most apparent role in getting the description of options right in choice situations involving uncertainty.[26] Revealed preference theory under risk and uncertainty either starts by describing agents as choosing between lotteries, that is, probability distributions over outcomes,[27] or as choosing between acts, that is assignments of outcomes to states of the world.[28] In these applications, according to the theory of options sketched in the last section, the specification of probabilities of outcomes, states of the world and assignments of outcomes to the states of the world must be consistent with the agent's beliefs about the choice situation in order for the revealed preference approach to identify stable and consistent preferences.

Interestingly, revealed preference theorists working on the theoretical foundations of the theory under risk and uncertainty, at least, appear to be happy to accept this. Kim (1996: 464), who applies revealed preference theory to the choice of lotteries, describes his approach as follows: "[T]he uncertainty modeled in this paper is limited to the situation where both the outside observer and the decision-maker agree on the probability distributions of the lotteries considered." It is hard to see how agreement on probabilities, where these are not themselves explicitly derived from choice behaviour, could be understood in non-mentalistic terms. It seems we must think of it either as an agreement in degrees of

[26] Getting the description of options right is also going to be tricky, and a substantive part of the analysis, in game theoretic applications of revealed preference methodology. However, revealed preference methodology is much less commonly applied in game theoretic settings, despite the existence of relevant representation theorems (see footnote 9 above).

[27] See, for instance, Green and Srivastava (1986), Border (1992) and Kim (1996).

[28] See, for instance, Bossert and Suzumura (2012) and Echenique and Saito (2015).

belief, or as agreement in judgement about objective probabilities. This phrasing also acknowledges the kinds of problems we have discussed: The success of the approach depends on economists accurately capturing how the options present themselves to the agent. More generally, when probabilities are presupposed by a revealed preference theorist working within the von Neumann-Morgenstern framework, probabilities are usually described to be 'known'. In the spirit of the above quotation, I propose this should be understood either as saying that objective probabilities should be known to both the agent and the economist,[29] or that the economist independently knows the agent's subjective probabilities.[30] Either way, assumptions about the agent's beliefs are made. In other applications, such as Green and Osband (1991), probabilities are treated as measuring epistemic states that reasonably respond to evidence. However, crucially, all these authors still insist on a behavioural interpretation of preference.

Empirical literature that aims to estimate utility or demand functions from choice data mostly[31] models consumption choices as choices under certainty: Agents are taken to directly choose some outcome, or consumption bundle. And so reference to probabilistic belief does not come up there. The problem that description of commodity bundles needs to be consistent with agents' beliefs about them is however still relevant here, and is admittedly little discussed in the core revealed preference literature. I propose that one major reason for this is that in the case of market commodities, the market to a large extent defines commodities: It demarcates commodities from each other, and presents consumers with a carefully curated package of information about the product. Since producers have an interest in repeat business, the kinds of misunderstandings involved in our wasabi case are going to be rare. And learning through repeat purchases is likely to quickly eliminate such anomalies. Widespread misinformation about commodities is going to be most likely in cases where commodities have long-term harmful effects that are either yet unknown or hidden by producers. But in these kinds of circumstances, economists will often be in the same epistemic situation as consumers. And if they are not, I submit that they would in fact bring the appropriate caution to revealed preference approaches. No economist would conclude, e.g., from data on cigarette consumption in the 1950s that consumers have a revealed preference for carcinogenic substances.[32]

[29] Chambers and Echenique's (2016: 114) textbook, for instance, understands probabilities within the von Neumann-Morgenstern framework in this way: "There are times when probabilities can be thought to be objective and known, or observable. This is the case, for example, when outcomes are randomized according to some known physical device – such as a game in a casino, or a randomization device used by an experimenter in the laboratory."

[30] This seems to be the way in which Green and Srivastava (1986) think of their framework, as they claim probabilities are both observable and subjective.

[31] The exception are empirical studies of the demand for insurance or betting behaviour. See, e.g., Jullien and Salanié (2000), Cohen and Einav (2007), and Barseghyan et al. (2013).

[32] Cases where there are genuine doubts about the true nature of a product amongst consumers and economists, as is the case, e.g. in markets for used cars as analysed in Akerlof's (1970) famous model, are usually analysed within the economics of information, which assumes a framework of uncertainty. As just

In the literature on the valuation of goods that are not directly traded in markets, such as environmental goods, discussions about "commodity definition" are, on the other hand, very common. And indeed, part of what is acknowledged to make commodity definition in this area difficult is that agents often lack information on how environmental harms and benefits are brought about. One way in which revealed preference methods are employed in this literature is in trying to infer the value of a good not directly traded on the market from the market value of related market commodities. To take an example from Boyd and Krupnick (2009), economists may want to infer the value of wetlands to the inhabitants of an area from the premium paid for houses in the vicinity of the wetlands. However, house buyers may not know of all the environmental benefits of the wetlands. This is one of the reasons why Boyd and Krupnick argue that we may be able to infer from the premium paid a revealed preference for proximity to open spaces, but not a revealed preference for the abundance of wildlife and clean water. Or, in other words, supposing that the description of the options house-buyers are choosing includes a description of those environmental benefits would be a misspecification: This description of the options is not consistent with the agents' relevant beliefs.

Related empirical methods include the use of contingent valuation surveys, where subjects are asked about what they would be willing to pay for various environmental goods. In this literature, it is generally acknowledged that subjects must be given as much information as possible about the environmental goods.[33] Again, this seems to be motivated by the thought that in order to legitimately elicit preferences over the options under the description we are actually interested in, we must ensure that that description of the options is consistent with the agents' relevant beliefs.

According to the second part of my proposed rule for the specification of options, economists should try to include in their description of options everything that makes a significant difference to agents' choice behaviour. Does this cohere with economic practice? Ultimately, the core interest of revealed preference methodology is the determination of demand functions and Engel curves for various commodities, which can then be used to predict and explain market behaviours and guide policy-making. The question of the appropriate description of outcomes, in this context, is the question of when commodities should be explicitly distinguished from each other. Along the lines I just suggested, the revealed preference theorist should make this decision by asking herself whether two commodities are different in a way that may affect agents' choice behaviour in the kinds of contexts in which we aim to predict and explain, namely, whether market demand is going to behave significantly differently for the two commodities. Indeed, these seem to be

discussed, in such a framework, options need to be explicitly described so as to capture agent's beliefs about the nature of the uncertainty they face. Note, however, that in Akerlof's core model, there is ultimately no doubt about the nature of the cars that are traded – they will all be bad.

[33] See, e.g., Carson (1998) on the valuation of tropical rainforests.

exactly the kinds of considerations driving the ways in which commodities are described in the empirical literature.

Many revealed preference studies utilize national expenditure surveys which generate data on household expenditures on various commodity groups. Blundell et al. (2003), for instance, use data from the British Family Expenditure Survey from 1974 to 1993, grouped into 22 commodity groups, including beer, wine, spirits, leisure goods and leisure services. The question of whether such categories are too coarse-grained is discussed extensively in the empirical economics literature as the question of 'disaggregation'.[34] Disaggregation is generally taken to allow for more accurate forecasts when a commodity group is disaggregated into smaller commodity groups for which consumer demand behaves significantly differently. But there may be practical limitations to the extent to which we can do so, and it is generally acknowledged that the desirable level of disaggregation depends on the purpose of the exercise, that is, whether we wish to predict and explain economy- or industry-wide phenomena, or rather movements within specific industries.[35] Suppose, for instance, we grouped all alcoholic beverages together, and estimated that share of expenditure spent on alcoholic beverages decreases with rising income. While this may accurately predict industry-wide phenomena as incomes rise, the prediction will not hold for all alcoholic beverages, with quality wines, spirits and craft brews potential exceptions.

How do we know that consumer demand behaves significantly differently for two commodities? Speculation about consumers' desires or mentalistic preferences may give us some indication here. Intuitively, it makes sense that people's demand for wine should react differently to changes in price and income than consumer's demand for beer. However, decisions about the right level of disaggregation in revealed preference studies on consumer demand are usually driven by more empirical considerations. For instance, there are many studies suggesting that demand for beer, wine and spirits respectively behaves quite differently in most countries.[36] For some commodity groups, differential demand behaviour may not be quite as intuitively obvious. For instance, it may be not as clear why demand for leisure goods and leisure services should behave differently. Here, evidence on choice behaviour will be especially helpful: Where the data has been disaggregated in the past, were significant differences found? If so, and unless we have good reason to think that conditions are different in the context of a present study, we have good reason to treat two commodities as different.

Of course, evidence of people's mental states, for instance from surveys or neuroeconomics, may also be helpful in determining what features of commodities make a difference

[34]For an edited volume dedicated to the issue, see Barker and Pesaran (1990).
[35]See, e.g., Barker and Pesaran's introduction in Barker and Pesaran (1990).
[36]See Fogarty (2010) for a review of the literature.

to consumption choices. In fact, more generally, non-choice evidence about agents' mental states is not irrelevant to revealed preference theory as I characterize it here. For one, as noted above, non-choice evidence of *belief* may be relevant in getting the description of options right. However, it is important to note that for the revealed preference theorist, mentalist evidence that agents desire certain features of options is less direct evidence than evidence of differential choice behaviour. It requires us to not only make the fallible inference from one context to another, but also to make the fallible inference from the presence of a desire relating to some feature of agents' options to differential choice behaviour. And so where there is ample choice evidence available, this will be the more informative evidence for the revealed preference theorist.

I have argued that, in order to identify an at least potentially stable and consistent preference relation from an agent's choice behaviour, economists must describe the options an agent is facing in a way that is consistent with her beliefs. Moreover, they must aim to include in the description of options any factors that they have reason to believe significantly affect the agent's choice behaviour in the contexts of interest. This means a fallible inductive step is involved in inferring preferences from choice behaviour, as economists may fail to correctly characterize the agent's options. The success of this inference relies, amongst other things, on economists correctly identifying the agent's beliefs. What the foregoing examples of economic practice show, I think, is that economists working within the revealed preference framework are not in principle opposed to appealing to mental states in precisely this way. Indeed, in the areas where they are most relevant to getting the analysis right – preferences over non-market goods, and choice under uncertainty – beliefs are explicitly discussed. Importantly, however, none of this is taken by these economists to show that preferences themselves aren't to be understood behaviourally, and rightly so. We can combine a partly mentalistic theory of options of the kind I have described here with a behavioural interpretation of preference.

Revealed preference theorists typically take the behavioural interpretation of preference to apply more comprehensively than revealed preference methodology. In particular, preferences as they feature in theoretical economic models can also be given a behavioural interpretation. As noted above, the claim here is that positing expected utility maximizing agents in theoretical economic models amounts to positing theoretical entities that choose in a certain way. Drawing on the theory of options developed in the last section, now we can add that the choice of description of the options these theoretical entities are choosing between captures both what we take these theoretical entities to believe about their environments, and what is relevant to their choice. And so, while according to the behavioural interpretation, these models involve no assumptions about conative mental states like preference in the ordinary sense, they often do involve commitments about beliefs.[37] When options are described in a way that implies that agents in these models have

[37]In some parts of economics, these commitments can indeed be quite strong, for instance when game

(true) beliefs that real agents typically don't, this is a potentially worrying disanalogy between model and real world. In fact, there is great controversy over assumptions of perfect information or common knowledge of rationality as they are implemented in many economic models. If my analysis is correct, revealed preference theorists can join in this criticism, as it pertains to the minimal mentalism they must grant. The only criticism they will cast off as misguided is the kind of criticism that complains that real agents don't have conative mental states corresponding to the preferences and utilities featuring in these models. And they are arguably right to cast this off as misguided.[38]

The charge against revealed preference theory considered in this paper is that preference attribution is inextricably linked to the attribution of mental states, in particular beliefs, and that this makes revealed preference theory incoherent or unmotivated. By showing that the appeal to mental states required is quite minimal and consistent with actual practice in revealed preference methodology, I have argued that revealed preference theory is not incoherent. The next section will argue that acknowledging this minimal mentalism does not make revealed preference theory unmotivated either.

6 The Appeal of Revealed Preference Theory

As we have seen, a limited appeal to mental states along the lines of the account I have presented is recognized to be necessary in practice by economists working within the revealed preference framework, rhetoric about "mindless economics" notwithstanding.[39] Still, critics of revealed preference theory might hold that this recognition undermines the key motivations for revealed preference theory. For instance, having noted the problem that choice reveals preference only once we have identified what the objects of choice are, and that verbal communication would be one natural way of finding out what the objects of choice are, Bradley (2017: 60) asks, "if recourse must be had to verbal communication then why not simply ask the subjects what they prefer and dispense with the pretence of purely behavioural evidence?" More generally, if the only motivation for revealed preference was supposed to be strict behaviourism, that is, a general rejection of non-behavioural evidence

theoretic solution concepts are justified in terms of the expected utility maximization of each agent, as in epistemic game theory. Here, the description of options will have to be consistent with the agent's beliefs about other players' choices, which will depend on their strategic reasoning. While the behavioural interpretation of preference can be extended to cover epistemic game theory in this way, the advantages of the revealed preference approach discussed below, in terms of being less committal about the mental causes of choice, will be less prominent here.

[38] Clarke's (2016) argument is instructive here.

[39] See Gul and Pesendorfer (2008) as an example of this rhetoric. That Gul and Pesendorfer's rhetoric does not do full justice to economic practice is also evidenced by the fact that Chambers and Echenique (2016: xv), in the first comprehensive textbook for revealed preference theory, declare in their preface that revealed preference theory does not preclude the use of data other than choice data.

and the presupposition of mental states not reducible to choice behaviour, then revealed preference theory now looks to be unmotivated. If the key goal was to get rid of appeal to mental states altogether, then what we have shown is that this goal is unattainable. In this section, I aim to show that coupling a partly mentalistic theory of options with a behavioural interpretation of preference still guarantees most of the core advantages economists have attributed to revealed preference theory over more mentalistic views.

The originators of revealed preference theory, most prominently Paul Samuelson, as well as, for instance, Little (1949), were writing in the heyday of logical positivism and behaviourism, and no doubt influenced by those intellectual currents. Still, it is not entirely clear that even those early revealed preference theorists were thorough-going behaviourists. Ross (2011), for instance, is sceptical, arguing that the goal for early revealed preference theorists was to continue with a process of eliminating psychological foundations from economics that started earlier than behaviourism. Apart from scepticism about the reality and empirical measurability of mentalistic notions of utility and preference specifically (and not necessarily mental state attribution in general), this was motivated by a desire to study aggregate economic dynamics and to be able to "ignore idiosyncrasies of individual consumers." (221)

Later defences of revealed preference theory echo several of the motivations that were driving early revealed preference theorists, without presupposing thorough-going behaviourism or positivism. I take there to be four core motivations of contemporary revealed preference theory. The first three are advantages that arise from black-boxing, as far as possible, the mental attitudes that cause choices. They apply to revealed preference methodology narrowly construed, as well as to the behavioural interpretation of preference in other parts of positive economics. This black-boxing is attractive, first, because it allows economists to retain a clearer disciplinary boundary to psychology and related disciplines. This desire is apparent in the vigour in which Gul and Pesendorfer (2008), for instance, fend off criticism of standard economic theory from psychology and neuroscience. Second, black-boxing is attractive in the face of scepticism about the specific psychological processes that would be presupposed under more mentalistic interpretations of expected utility theory. And third, even if we weren't sceptical about those psychological processes correctly describing at least some agents, black-boxing can help expected utility theory achieve greater generality, as it could apply even to agents who make decisions differently. In fact, it has no problem extending the analysis to choices driven by addiction, behavioural cues or subliminal advertising, or even the behaviour of non-human animals, as has in fact been done.[40] All we need for the theory to be fruitfully applied is that agents respond

[40]See Becker and Murphy (1988) on "rational addiction", and Kagel et al. (1995) for applications to non-human animals. Angner (ming) reviews further such applications of the theory to the decision-making of less than fully rational and reflective agents.

consistently to what they believe or register about their environment.[41]

The final core motivation of contemporary revealed preference theory, focused more on revealed preference methodology more narrowly construed, is to establish, in some sense, a 'tight connection' between the main data available to economists, namely data about market choices and contingent choice data (where consumers are asked to report how they *would* choose under various hypothetical circumstances), and the theoretical constructs of their theories. That is, it is taken to be desirable to infer preference from choice as directly as possible, making as few auxiliary assumptions as possible. Again, this is a core theme of Gul and Pesendorfer's (2008) defence of revealed preference theory, and is echoed in Chambers and Echenique's (2016) textbook on revealed preference theory, who claim that the point of revealed preference theory is to establish what economic models say about economic data, and in particular choice data. While positivists traditionally thought that only concepts defined in terms of measurement operations are meaningful, we can also give the concern for a tight connection between preference and choice data an empiricist reading: It is taken to be desirable to eliminate a fallible inductive step about mental states when using choice data to predict and explain market movements. This seems to me to be more uncontroversially an advantage of revealed preference methodology.

Importantly for us, none of these four core motivations are crucially undermined by the acknowledgement of a partly mentalistic theory of options. The mentalistic theory of options requires economists to make assumptions about agents' beliefs. While this is not consistent with strict behaviourism, it is consistent with black-boxing most of the motivating factors that bring about choice. It does not require us, for instance, to take a stance on whether choices are ultimately caused by expectation of hedonic utility, by an all-things-considered judgement of choice-worthiness, by impulses triggered by our environment, or by application of some rule of thumb. All we need to take a stance on is the agent's beliefs about her options. Doing so does not require integrating economics with full psychological theories of how agents make choices, and thus allows a greater disciplinary separation. It is consistent with scepticism about mentalistic notions of utility and preference playing a role in what causes choice. And it preserves the generality of the theory that allows it to capture choice behaviour brought about by a variety of different psychological mechanisms.[42]

[41] In the first instance, I here have in mind the generality of the revealed preference and expected utility *frameworks*: The frameworks can be used to analyze the choice behaviours of many different kinds of agents, who make choices in different ways. This would be an advantage even if the specific preferences we ascribe to different kinds of agents are all very idiosyncratic, because their choice behaviours look very different. Where they are not, and there are broad similarities in the choice behaviours of different agents, there is also the generality we get from ascribing the same preferences to many different kinds of agents. I take it this is the explanatory advantage Guala sees in preference ascriptions when their causal basis is multiply realized. Also see Ross (2011) on the advantages of this kind of generality.

[42] This is not to say that it might not be desirable to open the black box in some cases, in particular when there are systematic violations of the theory. In fact, this is a core motivation behind behavioural

What about the fourth core motivation? Even with a partly mentalistic theory of options, revealed preference theory preserves a tight connection between the data available to economists and their theoretical constructs. Granted, acknowledging a mentalistic theory of options means that preferences cannot be directly inferred from observation of choice behaviour, as economists must make fallible assumptions about beliefs. However, the inductive step from observation of behaviour to the existence of a preference is still safer under revealed preference theory than it is on most mentalist accounts. On the account presented here, the economist may make a mistake in her characterization of the agent's options. But if she doesn't, her inference from observed choice behaviour to preference is going to be correct. On the alternative picture, on the other hand, economists not only have to get the characterization of the choice problem right. They also make an inference from choice to whatever mental attitude preference is interpreted to be. Unless we subscribe to the functionalist kind of mentalism about preference described above (which I believe to be misguided), this further inference is going to be fallible.

7 Conclusion

By characterizing revealed preference theory as tied to a strict kind of (global or local) behaviourism, and then showing this behaviourism to be untenable, critics have been uncharitable towards revealed preference theory. Commitment to a behavioural interpretation of preference, which identifies preference with choice, is central to revealed preference theory. Beyond that, revealed preference theory is characterized by a methodological programme that aims to (i) use past choice data to ascribe preferences and utilities to agents, (ii) use these preference and utility ascriptions to make predictions about unobserved choices, and (iii) lay the formal foundations for the mapping of choice behaviour to preference. Under a behavioural interpretation of preference, this methodological programme has the advantage of largely black-boxing the causes of choice, thereby keeping a clearer disciplinary boundary to the psychological sciences, accommodating scepticism about mentalistic notions of utility and preference, and preserving greater generality of the theory. Moreover, it eliminates fallible inductive steps about agents' mental states when making predictions about future choice behaviour on the basis of past choice behaviour.

This paper has shown that revealed preference theory must and does take some mental states, namely beliefs, into account. However, I have argued that the extent to which it has to do so is fairly limited: It is restricted to taking beliefs into account in the description of the options an agent is choosing between. Such a partly mentalistic theory of options is not only consistent with retaining a behavioural interpretation of preference. Economic practice within the revealed preference framework is also by and large consistent with

economics.

pairing a partly mentalistic interpretation of options with a behavioural interpretation of preference. Moreover, the main motivations economists have cited for the behavioural interpretation of preference and revealed preference methodology are not undermined by adopting a partly mentalistic theory of options.

Of course, there may be reasons to question the standard motivations for revealed preference theory I have presented here. Moreover, there may be independent reasons to favour a mentalistic interpretation of preference over a behavioural one. For instance, one might think that, only under a mentalistic interpretation of preference can we view expected utility theory as providing rationalizing or folk-psychological kinds of explanations, or as being action-guiding.[43] I believe that these arguments ultimately fail to provide a justification for adopting a mentalistic interpretation of preference, and that the ability to black-box multifarious psychological processes, at least, is a good reason to favour the behavioural interpretation of preference. This is the subject of a separate paper. Here, I aimed to delineate more clearly where and how revealed preference theorists must, and indeed do, appeal to mental states, and moreover show that revealed preference theory should not be dismissed on the grounds that doing so renders the project incoherent or unmotivated.

Acknowledgments

I received generous comments on earlier drafts of this paper from Richard Bradley, Susanne Burri, Thomas Seiler, Roberto Fumagalli, participants of Gregor Betz's, Rafaela Hillerbrand's and Christian Seidel's research seminar at the Karlsruhe Institute of Technology, and three anonymous referees, which greatly improved the paper. I am also grateful for very helpful discussion when I presented the paper at the British Society for the Philosophy of Science Annual Conference in Oxford, and the Joint European Network for Philosophy of Social Science/Philosophy of Social Science Roundtable Conference in Hannover, both in 2018.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review*, 8(1):67–77.

Afriat, S. N. (1973). On a system of inequalities in demand analysis: An extension of the classical method. *International Economic Review*, 14(460-472).

[43]Indeed, Hausman's (2000) critique of a behavioural interpretation of game theory is at least in part based on the idea that game theory should be action-guiding.

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500.

Angner, E. (forthcoming). What preferences really are. *Philosophy of Science*.

Barker, T. and Pesaran, M., editors (1990). *Disaggregation in Econometric Modelling*. Routledge.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review*, 103(6):2499–2529.

Becker, G. S. and Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, 96(4):675–700.

Bermudez, J. L. (2009). *Decision Theory and Rationality*. Oxford University Press.

Bernheim, D. and Rangel, A. (2008). Choice-theoretic foundations for behavioral welfare economics. In Schotter, A. and Caplin, A., editors, *The Foundations of Positive and Normative Economics*, pages 155–192. Oxford University Press.

Binmore, K. (2008). *Rational Decisions*. Princeton University Press.

Blundell, R. W., Browning, M., and Crawford, I. A. (2003). Nonparametric engel curves and revealed preference. *Econometrica*, 71(1):205–240.

Border, K. C. (1992). Revealed preference, stochastic dominance, and the expected utility hypothesis. *Journal of Economic Theory*, 56(1):20–42.

Bossert, W. and Suzumura, K. (2012). Revealed preference and choice under uncertainty. *SERIEs*, 3(1-2):247–258.

Boyd, J. and Krupnick, A. (2009). The definition and choice of environmental commodities for nonmarket valuation. *Resources for the Future*, Discussion Papers dp-09-35.

Bradley, R. (2017). *Decision Theory with a Human Face*. Cambridge University Press.

Carson, R. T. (1998). Valuation of tropical rain forests: Philosophical and practical issues in the use of contingent valuation. *Ecological Economics*, 24:15–29.

Chambers, C. P. and Echenique, F. (2016). *Revealed Preference Theory*. Cambridge University Press.

Clarke, C. (2016). Preferences and positivist methodology in economics. *Philosophy of Science*, 83(2):192–212.

Cohen, A. and Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, 97(3):745–788.

Craver, C. and Alexandrova, A. (2008). No revolution necessary: Neural mechanisms for economics. *Economics and Philosophy*, 24:381–406.

Dietrich, F. and List, C. (2016). Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics and Philosophy*, 32(2):249–281.

Dreier, J. (1996). Rational preference: Decision theory as a theory of practical rationality. *Theory and Decision*, 40(3):249–276.

Echenique, F. and Saito, K. (2015). Savage in the market. *Econometrica*, 83(4):1467–1495.

Fogarty, J. (2010). The demand for beer, wine and spirits: A survey of the literature. *Journal of Economic Surveys*, 24(3):428–478.

Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge University Press.

Gilboa, I., Postlewaite, A., Samuelson, L., and Schmeidler, D. (2014). Economic models as analogies. *The Economic Journal*, 124(578).

Green, E. J. and Osband, K. (1991). A revealed preference theory of expected utility. *The Review of Economic Studies*, 58(4):677–695.

Green, R. C. and Srivastava, S. (1986). Expected utility maximization and demand behavior. *Journal of Economic Theory*, 38(2):313–323.

Guala, F. (2019). Preferences: Neither behavioural nor mental. *Economics and Philosophy*.

Gul, F. and Pesendorfer, W. (2008). The case for mindless economics. In Caplin, A. and Schotter, A., editors, *The Foundations of Positive and Normative Economics*. Oxford University Press.

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

Harrison, G. W. and Ross, D. (2018). Varieties of paternalism and the heterogeneity of utility structures. *Journal of Economic Methodology*, 25(1):42–67.

Harsanyi, J. (1977). On the rationale of the bayesian approach: comments on professor watkins's paper. In Butts, R. and Hintikka, J., editors, *Foundational Problems in the Special Sciences*. D. Reidel.

Hausman, D. (2000). Revealed preference, belief, and game theory. *Economics and Philosophy*, 16(1):99–115.

Hausman, D. (2012). *Preference, Value, Choice, and Welfare*. Cambridge University Press.

Houthakker, H. S. (1950). Revealed preference and the utility function. *Economia*, 17:159–174.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Jullien, B. and Salanié, B. (2000). Estimating preferences under risk: The case of racetrack bettors. *Journal of Political Economy*, 108(3).

Kagel, J. H., Battalio, R. C., and Green, L. (1995). *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge University Press.

Kim, T. (1996). Revealed preference theory on the choice of lotteries. *Journal of Mathematical Economics*, 26(4):463–477.

Little, I. (1949). A reformulation of the theory of consumer's behaviour. *Oxford Economic Papers*, 1(1):90–99.

Luce, R. D. and Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. Courier Corporation.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory*. Oxford University Press, 1 edition.

Pettit, P. (1991). Decision theory and folk psychology. In Bacharach, M. and Hurley, S., editors, *Foundations of Decision Theory: Issues and Advances*, pages 147–175. Blackwell.

Rosenberg, A. (1992). *Economics: Mathematical Politics or Science of Diminishing Returns?* University of Chicago Press.

Ross, D. (2011). Estranged parents and a schizophrenic child: choice in economics, psychology and neuroeconomics. *Journal of Economic Methodology*, 18(3):217–231.

Samuelson, P. (1938). A note on the pure theory of consumer's behaviour. *Econometrica*, 5(17):61–71.

Savage, L. (1972). *The Foundations of Statistics*. Wiley, second revised edition edition.

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7(1):1–31.

Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, 70(1):3–27.

Thoma, J. (2020). Folk psychology and the interpretation of decision theory. Unpublished manuscript.

Varian, H. (1982). The nonparametric approach to demand analysis. *Econometrica*, 50:945–974.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Biographical Information

Johanna Thoma is Assistant Professor in the Department of Philosophy, Logic and Scientific Method at the London School of Economics and Political Science. Her research interests lie in philosophy of economics, decision theory, practical rationality, ethics and public policy.