The Economic Concept of a Preference

Preferences have been central to economic theorizing since at least Adam Smith. According to Smith, market economies are founded in the human tendency to exchange in order to meet one's needs and wants. I need shelter but am a clumsy carpenter; you need food but are an impatient farmer. We decide to specialize, allowing me to buy your labor with the excess food I produce. Money is introduced, allowing me to buy your future labor with the produce I now sell. Consumer demand, founded in each person's preferences over combinations of goods, drives producers to allocate capital and labor to best meet that demand. And thus, the efficient market economy is born.¹

Preferences are thus descriptively central to economics. Agents are modeled as choosing according to their preferences – or, as if they so choose. Preferences are also normatively central: agents ought to choose so as to satisfy their preferences, and institutions ought to enable individuals to best satisfy their preferences.²

And yet, economists often seem to take preferences for granted, in at least three senses. The *content* of individual preferences is often taken for granted. Unlike psychology, economists often do not care why individuals have the preferences they do, only that they choose reliably and consistently according to whatever preferences they have. Data about individual choices is often taken to be the only data that bear on the building and testing of economic models and theories.³ Second, much of twentieth century economics is built on common sense assumptions about the *structure* of individual

¹ It is as if they are guided by an invisible hand to produce better outcomes for society (Smith 2004 [1776]: Book IV, Chapter 2).

² Pareto efficiency — that a distribution is morally good if and because no one can be made better off without making someone worse off, in light of their respective preferences — is the dominant ethical criterion for assessing institutions in economics. However, for early proponents of free markets such as Smith, markets are also good insofar as they promote freedom and egalitarian social relations Karl Marx famously criticized Smith's view of freedom and equality in the market as ignoring the domination of the laborer by the capitalist (1992 [1867]: Chapter 6). (Anderson 2017)

³ Dietrich and List (2016a: Section 3.2) and Gul and Pesendorfer (2008: Chapter 1) take one of the central stakes of the debate between neo-classical behaviorists and behavioral economists to be which evidence should be used to build and test models and theories.

preferences. For example, consumers are assumed to prefer consuming more of a good to less of it. Much of the focus of behavioral economics has been to build more psychologically accurate models by testing these assumptions.⁴ Finally, many areas of economics, such as macroeconomics, is concerned with the *aggregate economic patterns* that result from individual choices, rather than individual decision-making.

The concept of a preference thus occupies an uneasy position in economics. Preferences are central to the theories at the heart of economics, yet individual decision-making is often treated as if it is of little import. This uneasy position arises in part because of a longstanding disagreement in economics over whether preferences are *psychological* or *behavioral*. On one side are so-called mentalists, who argue that preferences are psychological. Mentalist interpretations trace their roots back to Smith and Jeremy Bentham, and are especially prominent among philosophers of economics, decision theorists, and behavioral economists. On the other side are so-called behaviorists, who argue that preferences are choice behavior. Behaviorist interpretations trace their roots back to late nineteenth and early twentieth century economists, such as Vilfredo Pareto and John Hicks,⁵ and is arguably still the dominant interpretation among economists.

This article addresses itself to the disagreement between mentalists and behaviorists over the concept of a preference. After introducing general background about the concept of a preference (Section 2), Section 3 discusses two desiderata on conceptions of a preference, one of which will be used to adjudicate the debate between behaviorists and mentalists. Sections 4 and 5 introduce mentalism and behaviorism respectively, and Section 6 lays out recent debates over whether

⁴ Much of the disagreement between behavioral economists and defenders of neo-classical economics is about the empirical status and content of key postulates(Guala 2019: 388). While this disagreement is scientifically important, it is orthogonal to the debate over different conceptions of a preference.

⁵ For example, Hicks (1956: 6, quoted from Sen 1982: 56) claimed that "econometric theory of demand does study human beings, but only as entities having certain patterns of market behavior; it makes no claim, no pretense, to be able to see inside their head."

behaviorism or mentalism is more predictive or explanatory. Section 7 concludes by discussing two potential paths forward.

2. Preferences

Despite vehement disagreement within economics, there is a core concept of a preference running through competing conceptions.⁶ That core concept is a *relative evaluation linked to choice*.⁷ Preferences are relative: agents have preferences for one object of choice over another. I prefer a beach vacation to a lake vacation, or a movie with a friend to a night out dancing. Preferences are also evaluative, or concern values, rather than facts. This is often cashed out in terms of the distinction between evaluative attitudes such as hope, desire, or disgust and cognitive attitudes such as belief, supposition, or doubt, although, as we will see, this way of doing so privileges mentalism over behaviorism. Finally, preferences are reliably connected with choices. On mentalist views, this is because preferences play a central role in practical reasoning about what to do. On behaviorist views, this is because preferences are identified with choices.

Preferences play a central role in the axiomatized choice theories that are the heart of standard microeconomics and some macroeconomics.^{8,9} Axiomatized choice theories include rational choice theory, decision theory, game theory, and social choice theory. These theories represent preferences as a binary relation whose relata are options, or mutually exclusive objects of choice. The formalization of preferences as a binary relation captures their relative nature, as one

⁶ Here I follow John Rawls' (1971) use of the distinction between a concept and conception to analyze theories of justice justice work on justice. I take there to be a shared concept of a preference across economics, picked out by the role it plays across these theories (representing agents' evaluative ranking of options). Different theories, however, may contain different interpretations of the concept, or conceptions.

⁷ Preferences are also sometimes represented numerically, as in expected utility theory. In economics, numerical preferences are standardly taken to reduce to relative preferences (Okasha 2016).

⁸ "Axiomatized choice theories" is taken from Herfeld (2018a). This article glosses over much of the diversity amongst axiomatized choice theories (Hands 2012, 2013; Herfeld 2018a).

⁹ Microeconomics since Alfred Marshall, as well as some macroeconomics, standardly models productive agents such as firms as profit maximizers and consuming agents such as individual or households as utility maximizers.

option is preferred to another. Individuals are also assumed to choose those options that are at least as good as the other feasible options in that choice context.¹⁰ This assumption links preferences to choice.

What about the evaluative nature of preferences? Other substantive assumptions about the preference relation determines its evaluative properties.¹¹ A standard assumption is transitivity – if the agent prefers x to y, and y to z, then she prefers x to z. Descriptively, agents are taken to have transitive preferences in a central class of cases. Furthermore, transitivity is often taken to be a normatively desirable property of an agent's preferences, indicating that she is consistent.

Axiomatized choice theories formalize the concept of a preference and embed it in theories that make further assumptions about preferences. Differing interpretations of this formalization give rise to competing conceptions. Sections 4 and 5 lay out the two most prominent types of interpretation: mentalist and behaviorist. But first, let's discuss two criteria to adjudicate the debate between these two conceptions.

3. Desiderata on the concept of a preference

Preferences play a central role in descriptive and normative theories of choice. This leads to the two criteria for judging candidate conceptions of a preference:

- (1) preferences are part of a plausible normative standard to assess choices, and
- (2) preferences predict and explain choices.¹²

Let's begin with criterion (1). As discussed in the introduction, preferences are part of normative standards for individual choices and institutions.¹³ Thus, any conception of a preference ought to be

¹⁰ Kreps (2013: 3).

¹¹ Other assumptions, by contrast, are made in order to make the formal apparatus tractable, such as the assumption of complete preferences (Bradley 2017).

¹² Adapted from Bermúdez (2009: Chapter 2). Bermúdez's proposed criteria are formulated for decision theory, and he includes a third criterion, that preferences structure deliberation and guide action.

¹³ Many economists would dispute this criterion and argue that economics is a not a normative science. For example, Gul and Pesendorfer (2008: 3) say that "Standard welfare economics functions as a part of positive economics."

able to play a role in these normative standards. However, for reasons of space, I will set criterion (1) aside, and focus on descriptive standards, in the form of criterion (2).

Criterion (2) inherits its plausibility from two general aims of all sciences, prediction and explanation. Sections 6 and 7 will further clarify the criterion, but an initial distinction will be helpful here. One might predict and explain choice from either a *first-personal* or a *third-personal* perspective.¹⁴ From a first-personal perspective, choosing what to do is based on reasons. Paradigm cases of choosing based on reasons are cases of deliberation informing action: I mull over the pros and cons of different dinner options, or compare job offers in Tokyo and Kyoto.¹⁵ More common in economics, however, is the explanation of choices from a third-personal perspective, in terms of causes. Such causal explanation does not require access to an agent's reasons. With enough data about your past choices, the theorist can predict and explain what you'll choose, with the aid of causal generalizations.

With that criterion in place, let's examine the two competing conceptions of a preference at issue: mentalism and behaviorism.

4. Preferences as mental

Aptly named, mentalist views take preferences to be evaluative mental states. Different mentalist views differ on what type of mental state preferences are. This section contrasts three prominent types of views: desire-based views, functionalist views, and total comparative evaluation views.

Interpretations of preferences as desires are standard in decision theory.¹⁶ Hedonic views dominated nineteenth century economics, and have been revived by recent research in neuro-

¹⁴ Bradley (2017), Kliemt (2009: Chapter 2).

¹⁵ Decision-theory, for example, models agential deliberation as the evaluation of acts in terms of how likely and desirable the possible consequences of an agent's actions are (see Bradley 2017: Chapter 1).

¹⁶ According to Dietrich and List (2013: 104), the standard interpretation of decision theory is "thoroughly Humean," representing agents as choosing so as to maximize their desires in light of their beliefs.

economics and behavioral economics.¹⁷ Different hedonic views may adopt a wider sense of desire as a judgment of choice-worthiness, or a narrower sense of desire as the expectation of pleasure.¹⁸ These views gain support from motivational assumptions linking preferences and choice. Both desire and preference exhibit a similar causal or conceptual connection between having a mental state and acting to make the world reflect that mental state — buying ice cream because I desire ice cream, or choosing udon because I prefer it to ramen.

In the late nineteenth century, interpretations of preference as desire were abandoned. Many of these theories made substantive commitments about the types of desires that motivated choice, which were argued to be predictively or explanatorily inadequate. In response, later theorists moved away from hedonic and cardinal notions of utility to focus on the ordinal preferences as ranking options. One way of interpreting this development is as the introduction of a more *functionalist* view of preference into economics, one that defined preferences by its causal role in producing choice via a ranking.¹⁹ Functionalism is a general theory of mind that individuates mental states by their functional role, or the causal relations in which they stand to other states, as well as the causal relations between inputs and outputs of the state.²⁰ On a functionalist view of preferences, a preference is any evaluative state that plays the right functional role. This functional role may include preferences' motivational roles, and causal relations between input perceptions of options and outputs of choices, as well as causal relations between preferences and other mental states like credences.

¹⁷ Hedonic views of "preference" in nineteenth century economics were rooted in Jeremy Bentham's psychological hedonism (Moscati 2018: Chapter 2). Modern research programs aim to measure the "true" utility, understood as a measurable psychophysical magnitude, that stands behind so-called decision utility, a mathematical representation of preference (Fumagalli 2019).

¹⁸ Schroeder (2015: 3.3).

¹⁹ As Hands (2014) discusses, earlier theoretical and empirical work in this revealed preference tradition by Pareto, Hicks, Samuelson, and others can be interpreted as mentalist.

Both desire-based and functionalist accounts are motivated by predictive and explanatory causal connections between preference and choice, in line with criterion (2). The final type of mentalist view we will consider aims to formulate a concept of preference that can predict and explain choice from both a first- and third-personal perspective. According to *total comparative evaluation* views, preferences are relative evaluations whose ranking reflects the agent's relevant reasons.²¹ However, from a third-personal perspective, agents' preferences need merely to be as if they were the result of the agent's considering of all the relevant reasons.²² Proponents argue that preferences as total comparative evaluations are better third-personal predictors and explainers of choice. For example, Hausman (2011: 35) argues that total comparative evaluations are more likely to be stable over time, or constant across the same choice contexts. That is because agents have not formed their preferences in other contexts on the basis of a different subset of reasons.

In sum, functionalist and total comparative evaluation views seem to be better third-personal predictors and explainers of choice, perhaps because the models in which they are embedded are more general. Subsequent sections will largely set aside these differences and discuss mentalist views generally, but such predictive and explanatory differences will be discussed where relevant.

5. Preferences as behavioral

The second family of interpretation we will consider are behavioral conceptions, which dominated much of twentieth century economics. Paul Samuelson's (1938) "A Note on the Pure Theory of Consumer's Behavior" marked an important milestone for behavioral interpretations. Samuelson showed analytically that the theory of consumer behavior can be modeled by demand functions — the amount of each good that will be purchased given a set of prices and the agent's income — and

²¹ Different authors differ as to what type of mental state a total comparative evaluation is. For example, Bradley (2017) argues that they are judgments, whereas Hausman (2011) takes them to be evaluative states that are subject to rational criticism.

²² Bradley 2017.

consistency constraints on those demand functions. The key consistency constraint proposed by Samuelson is the so-called the Weak Axiom of Revealed Preference, which states that if a consumer purchases a bundle of goods *B* at vector of prices *p* when bundle *C* is affordable, then, if *C* is chosen at prices *q*, then *B* is not affordable. The consumer's actions are taken to "reveal" a preference for *B* over *C*, due to her engagement in a certain pattern of consistent behavior.²³

Samuelson's contribution, and subsequent work in so-called revealed preference approaches to consumer demand, has been given two very different interpretations. On the first interpretation, a "revealed preference approach" to consumer choice is an inference procedure to identify an agent's psychological preferences from her choice behavior. The second interpretation, however, is the target of this article. On that interpretation, behavioral preferences represent behavioral, not psychological, information about an agent's choices.²⁴

A major choice point in fleshing out the behavioral conception is whether preferences represent actual or counterfactual behavior.²⁵ *Actualist* interpretations take preferences to summarize patterns in an agent's actual choice behavior. *Hypothetical* interpretations take preferences to summarize what agents would choose in a context, or a behavioral disposition to choose. The actualist interpretation is more in keeping with the some of the original empiricist motivations for work on revealed preferences.²⁶ However, it sits uneasily with much of economic practice, where

²³ Further theoretical work was done by Little (1949) and Houthakker (1950). Afriat (1967) then extended this theoretical work to the empirical study of consumer choice through the Generalized Axiom of Revealed Preference (Hands 2014; Kreps 2013: Chapter 4).

²⁴ As Clarke (2016) argues, a behaviorist interpretation of "preference" is separable from other theoretical commitments associated with traditional twentieth century behaviorism, such as the commitment that mental states do not exist. Many current behaviorists have jettisoned these theoretical commitments of traditional behaviorism: Binmore (2008), for example, acknowledges that mental states cause individual choices.

²⁵ Hausman (2011: 24-25).

²⁶ Hausman (2011) identifies Samuelson (1938) and Little (1950) as holding the actualist view, in light of their empiricist commitments, and Binmore (1994) as holding the hypothetical view.

economists use counterfactual information about agents' choices to infer what would happen.²⁷ Accordingly, this chapter focuses on the hypothetical behavioral interpretation, and drops the "hypothetical" qualifier henceforth.

6. Prediction and explanation

This section adjudicates the debate between mentalism and behaviorism in terms of criterion (2), the prediction and explanation of choice. To use criterion (2), I will further flesh it out in terms of causation and explanatory goodness.

Causation is central to prediction and explanation in economics. For example, one of the main aims of econometrics over the last three decades has been the estimation of causal effects across a wide variety of economic outcomes in different populations, which has become particularly important for the evaluation of policies. Causal models are often taken to be predictively and explanatorily superior to merely statistical models.²⁸ However, a causal model does not suffice for generating a good explanation of the phenomena; causal explanations of a single phenomenon can be better or worse than each other.²⁹ To adjudicate the debate, we should consider (1) whether behavioral and mental preferences cause choices (Section 6.1) and (2) whether they explain well (Section 6.2).

Even if behaviorist or mentalist preferences can be causes, however, theories that use one interpretation may be explanatorily superior to theories that use another. That is because causal

²⁷ Hausman (2011: Chapter 3, Section 2).

²⁸ For purposes of simplicity, I will assume that causal models are more predictive, although prediction and causality, as well as prediction and explanatory power, can and often do come apart (Scrivens 1962). Many social scientists assume that a more explanatory model is more predictive (Yarkoni and Westfall (2017) discuss the case of psychology).

²⁹ An example from Hitchcock and Woodward (2003: 184) is an explanation of plant growth in terms of an explanatory generalization relating water and fertilizer to plant height versus an explanations that appeals to plant physiology. While both use causal generalizations, the latter is, intuitively, better. Some philosophers of science also argue that there are non-causal explanations, which could also motivate the inclusion of Section 6.2 of this article, in addition to Section 6.1 (Lange 2016; Reutlinger and Saatsi (eds): 2018).

explanation does not suffice for good explanation. Section 6.2 considers arguments that either behaviorism or mentalism is explanatorily superior, due to being more general or cohesive.

6.1 Causation

This section examines whether one conception of preference is better suited to represent causal structure.³⁰ Mentalist interpretations are often taken to be superior: since mental states cause human actions, mental preferences belong to the right ontological category to qualify as a cause. Behavioral preferences, by contrast, seem ruled out from playing the role of a cause. How can an agent's past choices of coffee over tea cause her to choose coffee over tea today?

Most proponents of behavioral conceptions accept that behavioral preferences cannot be causes.³¹ They then argue that behavioral conceptions have other, weightier advantages. For example, some behaviorists are skeptical that a single concept of preference could be part of a general theory of decision-making that also causally explain decision-making, due to a wide heterogeneity in the causal mechanisms that produce choice.³² They take behavioral conceptions to be general but not causal, and mentalist conceptions to be causal but not general. Since axiomatic choice theories are a central part of the microfoundations of economics, generality ought to be privileged over causality.

However, there is room for proponents of behavioral interpretations to resist the claim that behavioral conceptions of a preference cannot be causes than proponents. To see this, it will be helpful to have an argument against behaviorism. I will focus on an argument that many

³⁰ The dialect of this section does not assume that the axiomatized choice models discussed in Section 2 represent the cause of choice behavior. In the case of standard axiomatized choice theories, there is a preponderance of empirical evidence that individuals do not choose as such theories represent them as choosing (see Camerer, Lowenstein, and Rabin (2004) and Thaler (1994) for evidence from behavioral economics). Furthermore, preferences may not be the only causes of choice, as axiomatized choice theories assume. Social and moral norms and expectations have been argued to determine behavior through psychological mechanisms cannot be accommodated by standard axiomatized choice frameworks (Cudd 2014; Pettit 1995; Sen 1982: Essay 4). For work that integrates norm-motivated behavior into non-standard formal choice models, see Bossert and Suzumura (2009), Bhattacharyya, Pattanaik, and Xu (2011), and Dietrich and List (2016b).

³¹ See Binmore (2008: 1.9) on what he calls the Causal Utility Fallacy.

³² Binmore 2008; Ross 2014b: 251-252.

philosophers have taken to be devastating for behaviorism.³³ It starts with a contrasting pair of cases, like the following:

Action Film. Zeynep prefers action films to comedies. She believes that there are action films and comedies playing in theaters in her city. She goes to see an action film.

Mistaken Action Film. Zeynep prefers comedies to action films. She believes that there are only action films playing in theaters in her city. She goes to see an action film.

Because of her mistaken belief, Zeynep chooses counter-preferentially. More generally, preference and choice can come apart, due to the causal influence of belief.³⁴ So, economist models must represent beliefs (premise 1). Beliefs are psychological states (premise 2). Furthermore, only psychological states can causally combine with other psychological states (premise 3). So, preferences must be psychological.³⁵

There is room for the behaviorist to challenge all three premises. I will sketch example challenges to each premise, starting with the first. Thoma (forthcoming) argues that the behaviorist has a principled alternative to dealing with problem cases like the above, one that does not require adding Zeynep's beliefs into the model to account for the case. The modeler can instead change the description of the options in the model. In *Mistaken Action Film*, Zeynep is better modeled as preferring to see an action film over no film, since seeing a comedy isn't compatible with her beliefs. More generally, the options ought to be picked out in a way that is consistent with the agent's beliefs and described in terms of features that matter for the agent's choice.³⁶ Since describing options in

³³ Hausman (2011), Rosenberg (1993), Guala (2019).

³⁴ Some economists, such as Gul and Pesendorfer (2008), seem willing to bite the bullet here.

³⁵ Many philosophers of economics take such cases to support a more devastating conclusion, that the conception of a behavioral preference is incoherent. That is because the conception seems to imply a contradiction: it claims that preferences are non-psychological, yet economists also must assume that preferences are psychological in models that utilize beliefs. Since the argument for this conclusion relies on the conclusion the argument as stated in the main text (preferences are psychological), I will focus on that argument.

³⁶ Thoma (forthcoming).

this way is desirable for choice modeling generally, the behaviorist has a principled response to supposedly devastating cases.

Some economics models, however, do explicitly or implicitly represent information. Perhaps there is a way to recover the anti-behaviorist argument by restricting the scope of premises 2 and 3 to economic models that represent information. However, some economic models that represent information cast premises two and three into doubt.

Consider premise 2, that beliefs are psychological states. In some models, such as models of markets with prices, the information represented may not be best interpreted as the psychological beliefs of particular agents. Markets are institutional environments that embed past information and channel agents' behavior without agents having to grasp the rules whereby the system works in their entirety.³⁷ One common way that markets do so is through prices, which can be modeled as signals that coordinate agents' actions by aggregating information without requiring agents to represent all the aggregated information in order to act on it.³⁸ Models of markets thus suggest that information is not always best modeled psychologically.

Now consider premise 3, that only psychological states can causally combine with other psychological states. An initial question is whether premise 3 is an *a priori* or an empirical truth. Since the latter is more in keeping with economics as an empirical science, we should look to the social sciences to see if there are examples of psychological beliefs and behavioral preferences jointly causing choice. Some game theoretic or agent-based models are plausibly interpreted as containing agents whose behavioral preferences and psychological beliefs cause their choices. Furthermore, practicing economists who are game theorists or work on agent-based models hold a behavioral

³⁷ Hayek (1948: Chapter 4), Ross (2005, 2014a), and Satz and Ferejohn (1994).

³⁸ See Milgrom (1979a, 1979b) and Wilson (1977) for results that in first price auctions with enough participants, the price aggregates information about common value. Well-functioning markets reduce private information about the uncertain characteristics of the good, thus avoiding problems raised by asymmetric information. The aggregating function of prices is some evidence that common information drive markets, not individual cognitive states.

conception, indicating that some economists take it as empirically plausible that behavioral preferences can causally combine with cognitive psychological states. In the very least, Premise 3 is not obvious, as long as it is understood as an empirical claim, not an *a priori* truth.

The above behaviorist challenges push back against the argument that behavioral preferences cannot be causes. If successful, this leave behaviorist and mentalist interpretations on a par as causal predictors of choice, for all that has been said.

6.2 Explanation

Why did mentalist conceptions seem so causally superior? One potential explanation is that mentalist conceptions are explanatorily superior to behavioral conceptions, even if they are not causally so. Explaining Abid's choice of coffee over tea in terms of his behavioral disposition to choose coffee over tea, says the mentalist critic, does inform us that he does so regularly, but leaves us not much more enlightened as to *why* he chooses coffee over tea. For that, we require an explanation of his behavior in terms of his mental states, such as his beliefs and desires.³⁹

To examine whether behavioral or mental conceptions have an explanatory advantage, I will focus on two criteria for a good explanation: generality and empirical similarity. First, generality. Interestingly, behavioral preferences are often taken to have an explanatory advantage when it comes to generality. This claim is often backed up by citing the inability of more psychologically realistic rival theories such as prospect theory to explain a wide range of choice behavior. Why might alternative theories fail in this way? The loss of psychological realism in standard axiomatized choice theories under a behavioral interpretation leads to more general explanations because individual psychology, as well as variation therein, is taken to be of negligible causal influence.⁴⁰ Instead, the

³⁹ Steele (2014).

⁴⁰ Dowding (2002), Ross (2005), and Guala (2019). In support of the general claim that variation in individual psychology does not influence standard results in microeconomics, one might appeal to attempts to found microeconomics in models without classically rational agents, such as Becker (1962) and Gode and Sunder (1993) (see Ross (2014b: Chapter 5) for discussion). Of course, variations individual psychology is sometimes causally significant in markets, as the success of behavioral finance has shown.

same small set of distinctly economic, aggregate level variables explain the outcomes of interest in a wide variety of systems. Thus, behavioral conceptions are explanatorily superior because they abstract away from causally irrelevant psychology, and are thus more general.

Note, however, that more work needs to be done to give behaviorism a conclusive explanatory advantage. Functionalist views, for example, abstract away from the psychological details of particular types of mental states. Furthermore, one might worry that behavioral interpretations lead to an inappropriate level of generality. Generality is not the only explanatory virtue. If it were, scientists could achieve spurious generality by adding irrelevant generalizations to their explanations and thereby improve them. But, intuitively, the explanation should be about the target system, i.e., explain in terms of generalizations that apply to an empirically similar range of phenomena.⁴¹

Behavioral explanations have been alleged to be spuriously general, i.e., too dissimilar. One criticism argues that behavioral preferences are so general that they are almost tautologous. An explanation of Abid's choice of coffee over tea in terms of his tendency to choose coffee over tea leaves us only slightly more enlightened as to why Abid chose coffee over tea than an explanation of why Gary is an unmarried man in terms of his bachelorhood. However, while behavioral preferences are highly abstract, they are not tautologous. Even under a behavioral interpretation, the binary preference relation summarizes information about regularities in individual behavior, regularities that are not entailed by the concept itself. Behavioral preferences are even more informative when embedded in a model: it is not trivial to figure out the logical consequences of ascribing a particular preference to individuals when modeling an individual decision problem or a system with interacting agents.

Perhaps behavioral preferences are not cohesive because they are not realistic enough. The charge that behavioral preferences are not realistic enough has been made forcefully by behavioral

⁴¹ Strevens (2008) call this "cohesion." The arguments here do not assume his or any other account of cohesion.

economists, who argue that the lack of realism makes for worse economics.⁴² The standard of realism is partly determined by other social sciences that study preferences, such as psychology, cognitive science, and neuroscience. Behaviorists, however, take issue with the claim that realism requires integration with other social sciences. Some are skeptical about a single standard of realism that applies across neuroscience, psychology, and economics.⁴³ One reason is a skepticism that a single social scientific discipline could formulate a general theory of human nature that can be used to explain phenomena of interest to many disciplines. Instead, each discipline uses abstractions suited to the phenomena it studies, and measures their empirical similarity by its own metrics.

An examination of the explanatory merits of behaviorism and mentalism has also produced a stalemate. While behavioral preferences are more general, functionalist interpretations of preference may be similarly abstract. The charge that behavioral conceptions are spuriously general has bottomed out in a disagreement about the boundaries of different social sciences and the independence of different theories in determining the meaning of their concepts. In the conclusion, we will examine two recent attempts to move the debate forward that pick up these two themes, respectively.

7. Conclusion: Beyond the behaviorism and mentalism divide?

Thus far, we have been treating mentalist and behaviorist views as mutually exclusive and exhaustive of possible conceptions of a preference suitable for economics. Perhaps, however, these conceptions are not exhaustive, and there is an alternative that inherits the advantage of both while avoiding their issues. Or, perhaps both conceptions can co-exist in economics in different theories. I will conclude by examining two recent contributions to the debate over the concept of a preference that take each of these approaches.

⁴² As Rabin (2002: 658) says, "ceteris paribus, the more realistic our assumptions about economic actors, the better our economics." See also Thaler (2000), Angner and Lowenstein (2012), and Hausman (2011).

⁴³ Gul and Pesendorfer 2008.

Guala (2019) argues for a conception of preference that is neither behavioral nor mental. He argues that preferences are belief-dependent dispositions.⁴⁴ Dispositions are tendencies to manifest some behavior under certain conditions. For example, salt is disposed to dissolve in water, or a glass to break when struck. In the case of preferences, they are dispositions to choose one option over another in a choice context.

One of the primary purported advantages of his view is that it can explain away the seeming multiplicity of conceptions of preference. Because axiomatized choice frameworks are successfully applied to very different kinds of agents, from human agents to hermit crabs to firms, one might take this empirical dissimilarity to support different conceptions of preference in economics.⁴⁵ Guala argues that his view allows a single concept to model these different agents, because dispositions can have multiply realizable causal bases. Dispositions' causal bases are the underlying properties that cause the object to manifest the relevant behavior under certain conditions. It is a fragile glass' microstructure, for example, that causes it to break when dropped. The causal basis of some dispositions are multiply realizable: the disposition to be in pain when struck may be differently physically realized in carbon and non-carbon brains.⁴⁶ Guala argues that the successful modeling of different physical kinds of agents by axiomatized choice theories lends support to the claim that the causal basis of a preference is multiply realizable. This multiple realizability may be argued to support a single metaphysical kind, or it may be argued to support a conceptual unity despite the lack of a single metaphysical kind.

This argument brings us back to the issue of empirical similarity. Guala's argument can be cast as one that empirical dissimilarity does not lead to conceptual disunity. More work needs to be

⁴⁴ Preferences must be belief-dependent, according to Guala, to incorporate the dependence of choice on belief.

⁴⁵ See also Satz and Ferejohn (1994). Elwood and Appel (2009), for example, attribute preferences to hermit crabs in order to explain when they leave different qualities of shells in response to electric shocks.

⁴⁶ Fodor (1974: 105-107) argues that entities with different underlying physical structures can realize the same state. As Block (1978: 264-265) notes in response to a point by Kim (1972) that any physical objects have *some* physical properties in common, the argument must be that there is no *non-trivial* physical kind or property that these realizers share.

done to argue that dispositional preferences are conceptually unified kind despite physical heterogeneity. Some *prima facie* evidence for its conceptual unity comes from the explanatory power of preferences across these dissimilar agents. Many philosophers of science, however, have thought that dispositions are explanatorily inferior to their causal bases. Molière's satire *Le Malade Imaginaire* famously mocks those who traffic in dispositional explanations. A group of doctors are quizzing an applicant to their professor, and ask the applicant why opium puts people to sleep. He answers "to this I reply that there is a dormitive virtue in it, whose nature it is to make the senses drowsy." Along with Molière, most philosophers of science judge this to be a poor explanation, especially when contrasted an explanation in terms of opium's biochemical mechanisms.⁴⁷ Like dormitive virtue, perhaps belief-dependent dispositions are so explanatorily inferior to their causal bases that they offer little or no explanation at all.

A second strategy for moving beyond the mentalism and behaviorism divide argues that there are multiple conceptions of a preference in economics. Here I will focus on a recent argument by Angner (2019), who draws on considerations about the context-dependent meaning of theoretical terms in science to argue for a pluralist view of preferences.

Theoretical terms are introduced into a scientific theory for a purpose that cannot be accomplished by previously well understood terms. Examples of theoretical terms include gravitational force in physics, genes in biology, or electrons in chemistry. The standard account of theoretical terms in philosophy treats them as entirely defined by the postulates of theory in which they are embedded, in terms of primitives or already well-defined terms.⁴⁸

⁴⁷ Molière (1926: volume viii, 328), quoted from Hutchinson (1991: 245). As Mackie (1977: 367-368) says, an explanation of why opium puts people to sleep that cites its dormitive virtue "merely provides a place-holder for the genuine explanation which would be given if the unknown property the causal process or mechanism by which it produces sleep were more explicitly described." Of course, not all philosophers of science are so critical of the doctors' approval (Lewis 1986).

⁴⁸ This characterization of the standard view follows Strevens (2012). An example view is one according to which theoretical terms are taken to be implicitly defined by correspondence rules, i.e., sentences that contain theoretical and non-theoretical terms and that thereby link theoretical terms to a number of non-theoretical terms in such a way as to generate an interpretation of those theoretical terms (Carnap 1939).

"Preference" is a strong candidate for being treated as a theoretical term, despite its prevalence outside scientific theories. In axiomatized choice theories, preferences are a theoretical building block defined by formal properties of preference rankings, such as transitivity and completeness in standard frameworks. These theoretical postulates determine the meaning of the concept, rather than its pre-theoretical use. If preference is indeed a theoretical term, then one would except disagreement over the concept to focus on the theory's postulates. Behavioral economics thus supports preference as a theoretical term: on of its aims is to test commonly assumed postulates and to develop new models of individual decision-making with different postulates that better fit the experimental data.⁴⁹

One might worry, however, that treating "preference" as a theoretical term leads to an implausible proliferation of conceptions. There may also be principled reasons to limit the plurality of conceptions of preference in economics, such as additional constraints on the meaning of theoretical terms. One proposed desideratum for choice theoretic models is that their concepts are continuous with those of folk psychology.⁵⁰ The mere addition of such constraints, however, neither entails a unified concept nor a uniform set of constraints across economics. For example, it may be that economics needs multiple concepts of "preference" that serve different theoretical purposes, some of which are better served by a continuity with folk psychology and some of which are not.⁵¹

Neither strategy for moving the debate forward is without its predictive or explanatory costs. More generally, both Guala's and Angner's views highlight the background disagreements that drive disagreements over conceptions of a preference, such as disagreements over the boundaries of

⁴⁹ Prospect theory, for example, was developed as an alternative to expected utility theory, on the basis of a large body of experimental work on risk attitudes in individual decision-making (Kahneman and Tversky 1979).

⁵⁰ Authors such as Lewis (1974), Hausman (2011), Steele (2014) take such a criterion to further support a mentalist interpretation of "preference." For pushback, see Thoma (forthcoming), who argues that decision theory under a behavioral interpretation better facilitates folk psychological explanations.

⁵¹ See Sen (1982: Chapter 4, especially 4.7) for what I take to be an example of this view.

social sciences or explanation using dispositions. To make progress, more attention ought to be paid to those background disagreements.

References

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

Anderson, E. (2017). *Private Government: How Employers Rule our Lives (and Why We Don't Talk About It)*. Princeton: Princeton University Press.

Angner, E. (2018). "What preferences really are." Philosophy of Science 84(4): 660-681.

Angner, E., and G. Loewenstein. (2012). "Behavioral economics." In U. Mäki (ed.), Handbook of the Philosophy of Science: Philosophy of Economics. Amsterdam: Elsevier, 641-690.

Aydinonat, N. Emrah. (2007). "Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation." *Journal of Economic Methodology* 14(4): 429-454.

Becker, G. (1962). "Irrational behavior and economic theory." Journal of Political Economy 70: 1-13.

Binmore, K. (1994). Game Theory and the Social Contract. Cambridge, MA: MIT University Press.

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

Bhattacharyya, A., P. K. Pattanaik and Y. Xu. (2011). "Choice, internal consistency and rationality." *Economics and Philosophy 27(2)*: 123-149.

Block, N. (1978). "Troubles with functionalism." In C.W. Savage (ed.), *Minnesota Studies in the Philosophy of Science Vol. IX*, Minneapolis: University of Minnesota Press, 261-325.

Bossert, W. and K. Suzumura. (2009). "External norms and rationality of choice." *Economics and Philosophy 25*: 139–152.

Bradley, R. (2017). Decision theory with a human face. Cambridge: Cambridge University Press.

Cairnes, J.E. (1872). "New theories in political economy." *Fortnightly Review* 17: 71-76. Reprinted in Peart, S. (ed), *W.S. Jevons: Critical Responses*, Volume 3. (2003). Abingdon: Taylor and Francis.

Camerer, C., G. Loewenstein, and M. Rabin (eds.). (2004). *Advances in Behavioral Economics*. Princeton: Princeton University Press.

Camerer, C., G. Loewenstein, and D. Prelec. (2005). "Neuroeconomics: how neuroscience can inform economics." *Journal of Economic Literature* 43(1): 9-64.

Carnap, R. (1939). Foundations of Logic and Mathematics, International Encyclopedia, vol. I, no. 3. Chicago: University of Chicago Press.

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

Cudd, A. (2014). "Commitment as motivation: Amartya Sen's theory of agency and the explanation of behavior." *Economics and Philosophy 30(1)*: 35-56.

Dietrich, F. and C. List. (2013). "A reason-based theory of rational choice." Noûs 47(1): 104-134.

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

Dietrich, F. and C. List. (2016b). "Reason-based choice and context-dependence: an explanatory framework." *Economics and Philosophy 32(2)*: 175-229.

Dowding, K. (2002). "Revealed preference and external reference." Rationality and Society 14: 259-284.

Elwood, R. and M. Appel. (2009). "Pain experience in hermit crabs?" Animal Behavior 77: 1243-1246.

Fodor, J. (1974). "The Special Sciences (or: the disunity of science as a working hypothesis)." *Synthese* 28(2): 97-115.

Fugamalli, R. (2019). "(F)utility exposed." Philosophy of Science 86(5): 955-966.

Guala, F. (2019). "Preferences: neither behavioral nor mental." *Economics and Philosophy 35(3)*: 383-401.

Gode, D., and S. Sunder. (1993). "Allocative efficiency of markets with zero-intelligence traders: markets as a partial substitute for individual rationality." *Journal of Political Economy* 101: 119-137.

Gul, F. and W. Pesendorfer. (2008). "The case for mindless economics." In A. Caplan and A. Schotter (eds), *The Foundation of Positive and Normative Economics*. New York: Oxford University Press, 3-39.

Hands, D.W. (2012). "Realism, commonsensibles, and economics: the case of contemporary revealed preference theory." In ed. A. Lehtinen, J. Kuorikoski, and P. Ylikoski, *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, Part II, Chapter 7. Abingdon: Routledge.

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

Hausman, D. (2011.) Preference, Value, Choice, and Welfare. Cambridge, UK: Cambridge University Press.

Hawthorne, J. O'Leary. (1994). "A Corrective to the Ramsey-Lewis Account of Theoretical Terms." *Analysis 54(2)*:105-110.

Hayek, F. von. (1948). "The use of knowledge in society." *Individualism and Economic Order*. Chicago: University of Chicago Press, Chapter 4.

Herfeld, Catherine. (2018a). "The diversity of rational choice theory: a review note." *Topoi* https://doi.org/10.1007/s11245-018-9588-7.

Herfeld, Catherine. (2018b). "Explaining patterns, not details: reevaluating rational choice models in light of their explananda." *Journal of Economic Methodology 25(2)*: 179-209.

Hicks, J.R. (1956). A Revision of Demand Theory. Oxford: Oxford University Press.

Hitchcock and Woodward. (2003b.) "Explanatory generalizations, Part II: plumbing explanatory depth." *Noûs 37(2)*: 181-199.

Hodgson, G. (2012). "On the limits of rational choice theory." *Economic Thought 1(1)*.

Houthakker H. (1950). "Revealed preference and the utility function." Economica 17: 159-174.

Hutchison, K. (1991). "Dormitive virtues, scholastic qualities, and the new philosophies." *History of Science 29*: 245-278.

Kahneman, D. and A. Tversky. (1979). "Prospect theory: an analysis of decision under risk." *Econometrica* 47(2): 263-291.

Kim, J. (1982). "Phenomenal properties, psychophysical laws, and the identity theory." *The Monist* 56(2): 177-192.

Kliemt, J. (2009). Philosophy and Economics 1: Methods and Models. Munich: R. Oldenbourg Verlag.

Kreps, D. (2013). Microeconomic Foundations I. Princeton: Princeton University Press.

Lange, M. (2016). Because Without Cause: Non-Causal Explanation in Science and Mathematics. Oxford:

Oxford University Press.

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

Lewis, D. (1983). "How to define theoretical terms." *Philosophical Papers: Volume I.* Oxford: Oxford University Press, 78-96.

Lewis, David. (1986). "Causal explanation." In David Lewis (ed.), *Philosophical Papers Vol. II*. Oxford: Oxford University Press, 214-240.

List, C. and P. Pettit. (2011). Group Agency. Oxford: Oxford University Press.

Little, I.M.D. (1949). "A reformulation of the theory of consumer's behaviour." Oxford Economic Papers 1.

Mackie, J.L. (1977). "Dispositions, grounds, and causes." Synthese 34: 361-370.

Mäki, U. (1990). "Scientific realism and Austrian explanation." Review of Political Economy 2: 310-44.

Mäki, U. (2001). "Explanatory unification: double and doubtful." *Philosophy of the Social Sciences 31(4)*: 488-506.

Marx, K. (1992 [1867]). Capital: Volume 1. New York: Penguin Random House.

Milgrom, P. (1979a). The Structure of Information in Competitive Bidding. New York: Garland Publishing Company.

Milgrom, P. (1979b). "A convergence theorem for competitive bidding with differential information." *Econometrica* 47: 679-688.

Molière, J.B. (1926). The Plays of Moliére. Trans A.R. Waller. Edinburgh.

Moscati, I. (2018). *Measuring Utility: from the Marginal Revolution to Behavioral Economics*. Oxford: Oxford University Press.

Nagel, E. (1961). The Structure of Science: Problems in the Logic of Scientific Explanation. London: Routledge & Kegan Paul.

Oshaka, S. (2016). "On the interpretation of decision theory." Economics and Philosophy 32: 409-433.

Pettit, P. (1995). "The virtual reality of homo economicus." The Monist 78(3): 308-329.

Rabin, M. (2002). "A perspective on psychology and economics." *European Economic Review* 46: 657-685.

Rawls, J. (1971). A Theory of Justice. Cambridge, MA: Belknap Press.

Reutlinger, A. and Saatsi, J. (eds). (2018). *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford: Oxford University Press.

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

Ross, D. (2005). *Economic Theory and Cognitive Science: Microexplanation*. Cambridge: MIT University Press.

Ross, D. (2014a). "Pyschological versus economic models of bounded rationality." *Journal of Economic Methodology 21(4)*: 411-427.

Ross, D. (2014b). Philosophy of Economics. London: Palgrave Macmillan.

Samuelson, P. (1938). "A note on the pure theory of consumer's behaviour." Economica 5: 61-71.

Samuelson, P. (1955). The Foundation of Economics. Cambridge: Harvard University Press.

Satz, D. and J. Ferejohn. (1994). "Rational choice and social theory." Journal of Philosophy 91: 71-87.

Sen, A. (1982). *Choice, Welfare, and Measurement.* Cambridge, MA: The MIT Press, Chapter 4: "Rational fools: a critique of the behaviorist foundations of economic theory."

Schroeder, T. (2015). Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), URL = https://plato.stanford.edu/archives/sum2017/entries/desire/.

Scriven, M. (1962). "Comments on Professor Grunbaum's Remarks at the Wesleyan Meeting." *Philosophy of Science 29(2)*: 171-174.

Steele, K. (2014). "Choice models." In N. Cartwright and E. Montuschi (eds.), *Philosophy of Social Science: A New Introduction*. Oxford: Oxford University Press, 185-207.

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

Strevens, M. (2008). Depth. Harvard: Harvard University Press.

Strevens, M. (2012). "Theoretical terms without analytic truths." Philosophical Studies 160(1):167-190.

Thaler, R. (1994). The Winner's Curse: Paradoxes and Anomalies of Economic Life. Princeton: Princeton University Press.

Thaler, R. (2000). "From homo economicus to homo sapiens." *Journal of Economic Perspectives 14(1)*: 133-141.

Thoma, J. forthcoming. "In defense of revealed preference theory." Economics and Philosophy.

Thoma, J.

Vredenburgh, K. (2020). "A unificationist defense of revealed preference." *Economics and Philosophy* 36(1): 149-169.

Weber, M. (1978 [1922]). "The nature of social action." Wirtschaft und Gesellschaft. Reprinted in W.G. Runciman (ed.), Max Weber: Selections in translation. Cambridge: Cambridge University Press, 59-77.

Weisberg, M. (2013). Simulation and Similarity: Using Models to Understand the World. Oxford: Oxford University Press.

Weslake, B. (2010). "Explanatory depth." Philosophy of Science 77(2): 273-294.

Wilson, R. (1977). "A bidding model of perfect competition." Review of Economics Studies 4: 511-518.

Yarkoni, T. and J. Westfall. (2017). "Choosing prediction over explanation in psychology: lessons from machine learning." *Perspectives on Psychological Science* 12(6): 1100-1122.