A Unificationist Defense of Revealed Preferences

Kate Vredenburgh
Harvard University
kvredenburgh@fas.harvard.edu

1. Introduction

There is, by now, an impressive collection of criticisms of classical rational choice theory. Behavioral economists criticize its highly idealized psychological assumptions; some social scientists criticize its methodological imperialism; philosophers criticize just about everything.

One of the most maligned versions of rational choice theory are so-called revealed preference approaches (terminology following Hands (2013)). Revealed preference approaches, which take choice behavior as primitive, originated in the work of Samuelson (1938) and are still prevalent in empirical research, economics textbooks, and methodological debates. This stability in the working methodology of microeconomics is *prima facie* surprising, considering the weight of theoretical challenges and the (arguably more devastating) empirical challenges to classical rational choice theory more generally.[1]

This paper focuses on explanatory challenges to revealed preference approaches. In the spirit of Ross (2005, 2014), I aim to address important theoretical challenges in a manner that explains the stability of microeconomic methodology. In particular, I construct a new explanatory defense of revealed preference approaches that aims to rebut two seemingly devastating explanatory criticisms.

---

[1] As Ross (2005: 215) puts it, "one crucial fact that cannot be ignored – must, indeed, be explained, as philosophers of economics such as Rosenberg (1992) and Hausman (1992) have explicitly recognized – is that the working methodology of microeconomics, and its associated basic pedagogy, has shown much greater stability than its high-level theoretical commitments."

As a way into these criticisms, consider the following dialogue between Janelle, an undergraduate who has recently decided to study economics, and her parent. In this scenario, Janelle, enamored with her new course of study, has decided to explain her choice to study economics using revealed preference theory.

> JANELLE: I've decided to study economics!
>
> PARENT: Wonderful! Why did you choose economics?
>
> JANELLE: Well, I prefer it to the other available courses of study.
>
> PARENT: That's great, dear. I'm so glad that you found a course of study that both makes you happy and builds a widely applicable skill set.
>
> JANELLE: Ah. I think that you're confused about what I mean by "prefer." Statements about my preferences are just summaries of my choice behavior.
>
> PARENT: (becoming frustrated) You're starting to go in circles, Janelle. Let's start over: why did you choose economics as your major?
>
> JANELLE: Sorry, but I just don't get the question.

This dialogue brings out how misleading the label "revealed preferences" is. One might think that revealed preference theorists use choice behavior to get at, or reveal, the psychological preferences that cause that behavior. However, as Janelle's insistence on staying at the level of choice behavior illustrates, no preferences – in the psychological sense of preferences – are revealed. Instead, revealed preferences are a convenient way to talk about choice behavior, as detailed in Section 2.

This characterization of revealed preferences leads to two seemingly devastating explanatory problems. The first problem is that revealed preference explanations are circular. Janelle begins by stating that she chose economics because she prefers it to other available courses of study at her university. She then states that the (revealed) preference relation is nothing more than a summary of

patterns in choice behavior. So, Janelle has invoked a pattern among her choices to explain a particular choice, which seems to be devastatingly circular. I will discuss this so-called *problem of self-explanation* in Section 3.

There seems, though, to be an even more basic explanatory problem with Janelle's responses, one encapsulated by Janelle's answer to the question of *why* she chose economics. Janelle's answers don't seem to explain her choices at all, if a causal theory of explanation is assumed. This problem also appears to be a serious one: All else equal, a theory that explains the phenomena of interest is better than one that does not. I will discuss *the problem of causal explanation* in Section 4.

Thus far in the literature, the reaction to these two problems, among both proponents and critics of revealed preference approaches alike, has been to accept that the explanatory game is up for revealed preference approaches. That acceptance, though, has been too quick. In Sections 5, I argue for a new line of response to the explanatory criticisms. There, I argue that proponents of revealed preference approaches should adopt a different background conception of explanation from their critics, *unificationism.* Thus, the criticisms of Sections 3 and 4 miss the mark, because they assume a background conception of explanation that the revealed preference theorist should not – and, indeed, does not – hold.

The unificationist defense of revealed preference approaches given in this paper is, it must be said, a limited one. Unlike Gul and Pesendorfer (2008), Bernheim and Rangel (2009), or Binmore (2009), I do not make the case for revealed preference approaches over other approaches to modeling choice behavior, nor do I make the case that revealed preference approaches are more explanatory – in either the unificationist or the interventionist sense of "explanation" – than frameworks that represent psychologically real preferences. Instead, I give a conditional argument: if

revealed preference approaches are combined with unificationism about explanation, then these frameworks can escape the explanatory problems posed in Sections 3 and 4.

The conclusion discusses two interesting upshots of Sections 3 through 5. First, the conditional defense points towards a promising non-conditional defense of revealed preference approaches: the development of an epistemic theory of explanation, such as unificationism, paired with an argument that revealed preference approaches are more explanatory precisely because they abstract away from the causal role of preferences. In other words, the paper points to how one might develop a more epistemic or instrumentalist account of the explanatory role of preferences in economics and the social sciences, as a response to realist criticisms thereof, such as Dietrich and List (2016a).

Second, the paper opens up space for a re-examination of the view that revealed preferences do not causally explain. Section 4 argues that interventionism about explanation, paired with an interpretation of revealed preferences as behavioral dispositions, seems to vindicate the causal explanatory power of revealed preferences. Thus, the arguments of this paper actually support two distinct new lines of defense of the explanatory power of revealed preference approaches, one based on unificationism, another based on interventionism.

2. Revealed preference approaches

Revealed preference approaches are a set of frameworks within the rational choice theoretic approach to modeling choice behavior. To set up the discussion of revealed preference approaches, I will first introduce some basic building blocks of rational choice theory and then discuss the revealed preference relation.

Within classical rational choice theory, there are two main approaches to modeling choice behavior. The first approach, which I will not discuss, takes preferences as primitive, and imposes certain rationality requirements on an individual's preferences. The second approach takes choice behavior as primitive, and imposes rationality requirements on choice behavior.[2]

To model choice behavior, the following are taken as primitive:

- A non-empty set X of mutually exclusive objects of choice. Elements include $x, y, z\ldots$

- A non-empty set $\mathcal{K}$ of contexts, where each element $K \in \mathcal{K}$ is a non-empty subset $K \subseteq X$ of feasible options. K is often taken to be the set of all non-empty subsets of X.[3]

- A choice function $C: K \in \mathcal{K} \longrightarrow C(K) \in \mathcal{K}$, where the choice function assigns a non-empty set of options to every context.

In addition to the above primitives, revealed preference approaches then define a *revealed preference relation*. This binary relation imposes additional structure through an ordinal ranking of alternatives across contexts. One benefit of this additional structure, which I will return to in Section 5 and in the conclusion, is that the ordinal ranking more efficiently systematizes the relevant pattern of choice behavior, compared to the choice function. Importantly for what follows, the revealed preference relation does not represent any additional information, beyond this ordinal ranking; in particular, it does not represent any information about causes. Instead, it more efficiently encodes the information in the choice function. By efficiently summarizing patterns in choice behavior, the

---

[2] See Mas-Collel, Whinston, and Green (1995: Chapter 1), as well as Kreps (2013: Chapter 1), and Dietrich and List (2016b). Economists disagree over which of these two approaches is standard. Mas-Collel, Whinston, and Green (1995: 5), authors of a standard graduate microeconomics textbook, claim that "This preference-based approach is the more traditional of the two." Binmore (2009: 20), by contrast, claims that the revealed preference approach is "the official doctrine of all neo-classical economics, enshrined in all respectable textbooks."

[3] In order to prove certain theoretical results, K is taken to be the set of all non-empty subsets of X. By contrast, the modeling of a particular agent's decision problem usually includes only some of the non-empty subsets of X.

revealed preference relation serves an epistemic function, as it provides a more efficient tool for organizing information about such patterns.

There are many ways to construct the revealed preference relation; the construction method does not matter for present purposes, so I will adopt the following definition for the so-called weak revealed preference relation "$\succsim^*$." As above, x and y are options, K is a context, and $C(K)$ is the choice function.

$x \succsim^* y$ iff for some $K, x \in C(K)$ and $y \in K$

In other words, x is weakly revealed preferred to y if and only if x is chosen from a context that contains y (note that this is compatible with y being chosen over x in a different context).[4]

The weak revealed preference relation is a formal object in a theory, one that needs to be interpreted. The interpretation under consideration here takes the revealed preference relation to be a mere summary of patterns in choice behavior.[5] Here, I follow Clarke (2016) in maintaining that the claim that preferences are summaries of an agent's choice behavior is separable from behaviorist and other positivist claims that have often been adopted by proponents of revealed preferences (going back to Samuelson (1938) and still prevalent among some practicing economists). As discussed in the introduction, the revealed preference relation, so interpreted, is inaptly named, since preferences,

---

[4] By contrast, if an agent strictly prefers x to y, then the agent would always choose x over y, in contexts with both x and y. Finally, if the agent is indifferent between x and y, then either x or y may be chosen in contexts with both x and y. Both the strict revealed preference relation and indifference can be defined from the weak revealed preference relation. In what follows, I will not distinguish between these relations, and will instead use the term "revealed preference relation" rather loosely, to apply to the weak revealed preference relation, the strict revealed preference relation, and indifference. That is because the explanatory objections and response apply to all three relations.

[5] There are other interpretations of the revealed preference relation that do not lead to the explanatory problems of Sections 3 and 4. Dietrich and List (2016a: 266), for example, discuss what they call the "epistemological revealed preference thesis", namely, the view that the evidential basis for an economic theory should be restricted to choice behavior. This epistemological interpretation fits well with the self-understanding of many economists, who often understand revealed preferences as a methodology to infer the value of unobservable parameters (preferences) from observable parameters (choice behavior in markets) (see Hausman (2012: 23--24) for discussion).

understood as mental states attributable to agents, are not represented by the revealed preference relation. Instead, the revealed preference relation is shorthand for information about choice behavior. For example, say that Monique is represented as having a weak revealed preference for pineapples over bananas. This revealed preference relation is shorthand for a longer claim, of the following form: "In context K_1, Monique chose pineapples over bananas; in context K_2, Monique chose pineapples over bananas..."

There are two areas of economics where revealed preference approaches have been most influential, and thus where the arguments of this paper have the most significant ramifications for economic practice. The first is general equilibrium theory, particularly the study of demand side of markets. The study of demand often proceeds by starting from individual demand functions and aggregating those functions into a market-wide demand function.[6] The second is an empirical research program that Wade Hands calls contemporary revealed preference theory, which studies choice behavior by constructing a utility function that rationalizes a set of finite choice data. The bulk of this work has been done in applied research on consumer choice, which uses data about prices and quantities and the associated opportunity sets (see Blundell (2005)), but "contemporary revealed preference theory" covers "those who are using it as the primary theoretical tool in their applied research" (Hands 2012: Chapter 7, Section 3).

---

[6] One could, of course, provide a different foundation for demand theory, as explored in Becker (1962), Gode and Sunder (1993) and Hildenbrand (1994) (see Ross (2014: Chapter 5) for discussion). More generally, economists often make macro-level predictions and explanatory claims that are neutral across different micro-level units. Thus, the discussion of this paper does not have direct implications for those neutral claims. However, the question still arises: what are the most basic units in the economy, and how do they interact? The explanatory power of different micro-theories can, and should, be used to answer this question. Thus, this is a contribution to the larger question of the best foundation for demand theory. (Note that one answer to the question of the foundations of demand theory is that no general micro-foundation can be given, perhaps for reasons of within-population heterogeneity (see Ross (2005: 257)). I take it that a plausible defense of this answer, however, will argue against different micro-foundations on explanatory and other grounds.)

A caveat is in order here. There is widespread disagreement about how to interpret economic practice here, made more difficult by the following three-fold distinction. There is what economists actually do: how they construct theories, build and test models, etc. There is also what they say they do: how they say they build and test models, etc. One difficult question facing those interested in the methodology of economics is the question of how much weight the latter should be given in interpreting the former. Complicating the interpretation of economic practice is the further question of what economists *should* do. This latter question, for example, is one that Hausman (2012) addresses in a recent book, arguing that economists should (and do) interpret preferences as total subjective comparative evaluations.

Many philosophers interested in the question of what economists actually do have interpreted economists as, by and large, working with a mentalist interpretation of preference.[7] Accordingly, they interpret contemporary revealed preference approaches as a methodology to discover mental preferences (see footnotes five and seven for examples). Here I agree with Angner (forthcoming) that a better interpretation of economic practice is one on which economists take the concept of preference as a primitive – as neither capable of or in need of further analysis – with two key properties, namely, completeness and transitivity.[8] At the same time, economists engaged in empirical research need to identify the entities in the world that are represented by their models. There, I take the hypothetical revealed preference interpretation, discussed in Section 4, as the

---

[7] For example, Guala (2012: Chapter 6, Section 4) states that: "the decline of behaviorism, together with sustained conceptual criticism, have tempered this radical interpretation [that identifies preferences with consistent patterns of observable choice] over the years to the point that nowadays 'revealed preference theory' for most economists is no more than an empiricist commitment to derive demand curves from choice data, or to construct models that can be tested against observable behavior."

[8] An alternate, nearby interpretation of economic practice is one where economists take it as definitional of the preference relation that it is complete and transitive.

prevailing interpretation used by some communities of economists engaged in empirical research (see Section 3 for discussion of contemporary revealed preference approaches). That being said, I am happy to grant that the mentalist interpretation of preference prevails in economics, for dialectical purposes. The arguments of this paper primarily address the question of how economists should interpret revealed preference frameworks, in light of the two explanatory objections discussed in Sections 3 and 4.

With that background in place, let's move on to the two explanatory objections.

3.   The no self-explanation objection

The criticism that the identification of preferences (or utility) with choice behavior produces a circularity goes back at least to J.E. Cairnes' (1872: 20) criticism of W.S. Jevons' discussion of so-called "exchange-value," i.e., marginal utility.[9] Indeed, the circularity seems so obvious as to need no argument: A revealed preference relation is defined in terms of choice behavior, and is then used to explain choice behavior.

A mere circularity, of course, might not worry the revealed preference theorist. Jevons, Sen, and others, however, have argued that the circularity is a vicious one, either because of issues around the measurement of utility (see Cairnes (1872: 20)) or because it trivializes the concept of utility (see Sen (1977: 322--323)). Thus far, the problems that circularity raises for measurement and for the meaningfulness of economic theory have received the most attention in the literature.

---

[9] Revealed preference approaches were, of course, not on the scene when Cairnes was writing – Cairnes' direct target was Jevons' proposal both to measure the marginal utility of a commodity in terms of the price thereof and to explain price in terms of marginal utility.

As I will argue below, this circularity also raises a distinctly explanatory problem and thereby seems unavoidably vicious. Because of the circularity, explanations containing the revealed preference relation violate a popular and plausible principle of explanation, the *No Self-Explanation Principle* (taken from Miller (2015)). The No Self-Explanation Principle says that no fact (partially) explains itself. If this principle doesn't immediately strike you as plausible, think about a putative explanation of why a projectile travelled 10 meters that appeals, among other facts, to the fact that it travelled 10 meters, and about whether you would find that explanation satisfying.

Unfortunately, as I will argue below, this violation of the No Self-Explanation Principle is needed to deal with a particularly gnarly version of the problem of induction. Furthermore, an attempt to avoid the circularity yet deal with this version of the problem of induction escape the circularity on in letter, not spirit.

With that motivation in place, let's look more closely at the circularity. Here, it will help to have an example purported explanation. Say that a revealed preference theorist is trying to explain why Olivia chose olive pants over jeans or yellow pants. The revealed preference theorist would say that Olivia has a revealed preference for olive pants over jeans or yellow pants. So, the constructed revealed preference relation for Olivia predicts, and thereby explains, that she chooses the olive pants.[10] However, recall that the revealed preference relation is a mere summary of patterns among choices. So, information about choice behavior is used to explain a particular choice.

---

[10] Note that the revealed preference theorist has already made an unusual explanatory commitment here by identifying prediction and explanation. Scriven (1962: 54), criticizing Hempel and Oppenheim's (1948) deductive nomological account of explanation, observed that successful prediction and successful explanation often come apart – cows lying down in a field may be a reliable predictor of rain, but no explanation thereof. Since such criticisms, theorists have taken explanation and prediction as importantly different. The close identification of explanation with prediction by the rational choice theorist signals her commitment to a different background view of explanation, and fits better with the unificationist discussion of Section 5 than with the attribution of a causal account.

So far, it seems, so good: it is good scientific practice to use information about patterns among phenomena of some type to explain token behavior of that type. For example, an explanation of why Taib's ice cream melted on a hot day may give information about the melting point of ice cream, heat transfer between the ice cream and the surrounding air, etc. – information that applies to many token melting ice creams. However, grouping together thermodynamic and revealed preference explanations may strike the critic as slightly disingenuous. Taib's melting ice cream is not explained by mere reference to the fact that past ice cream cones have melted under similar conditions. Instead, the explanation cites robust generalizations that apply to a range of objects. In the case of revealed preference explanations, an agent's past choices of an option explain her future choices of that very same option (in conjunction with additional information about her budget, physical constraints she is under, etc.), rather than any appeal to more widely applicable generalizations.

Even the fact that choice behavior explains choice behavior, however, does not yet lead to circularity. Circularity only threatens if the particular choice to be explained is contained within the revealed preference relation. If the revealed preference relation includes the choice itself, then a particular choice partially explains itself.[11] Thus, a closer look at the domain of the revealed preference relation is an important next step. Following Hands (2013), I will divide revealed preference approaches into two major camps, each of which posits a different domain: more

---

[11] The circularity also arises for explanations of patterns of choices using a revealed preference relation. Say, for example, that individuals in a market are represented as having identical utility functions, and that fact is used to derive the aggregate demand function for some good. In other words, a set of token patterns (individual demand functions) are fit into a general pattern (an aggregate demand function). The aggregate demand function is then used to explain facts about consumption choices in the market. That aggregate demand function, which is a summary of token patterns of choice behavior, is thereby used to explain those same token patterns of choice behavior. Ergo, circularity.

theoretical approaches, and contemporary revealed preference approaches, an empirical research program.

The more theoretical tradition of revealed preference frameworks, one that goes back to Samuelson (1938), assumes an infinite domain of choice acts for the revealed preference relation. In such frameworks, the circularity is straightforward, as the domain of the revealed preference relation includes the choice that is explained thereby.

By contrast, contemporary revealed preference approaches focus on the construction of a revealed preference relation on the basis of a finite set of observed choice data (see Afriat (1967), Varian (1982), as well as an overview in Hands (2013: Section 2)). Such empirical approaches seem to posit a revealed preference relation with a domain of actual, observed choices.[12] In that case, there is no circularity, since the finite set of observed choice data does not contain the choice to be explained.

Such a solution, however, avoids an explanatory circularity by running up against a particularly serious version of the problem of induction. The restriction of the domain of the revealed preference relation to past choices raises the following inductive puzzle: why are we justified in using information about past choices to predict future ones? If the revealed preference relation were understood as a mere summary of past choices, it is not clear why past behavior would be a good guide to future behavior. Absent certain assumptions, my choice of a cappuccino this morning, for example, is not a good guide to my beverage choices of the afternoon. (Perhaps I am a caffeine sensitive person who only drinks coffee in the morning, for example.)

---

[12] As Gul and Pesendorfer (2008: 8) say, "standard (positive) theory identifies choice parameters from past behavior and relates these parameters to future behavior."

At this point, you might be thinking something like the following: "so what? The problem of induction is everyone's problem." Agreed – but, the problem of induction is particularly pressing in the case of contemporary revealed preference approaches. That is because, as Hands (2012: Chapter 6, Section 4) says:

> the natural inclination is to think that CRPT [contemporary revealed preference theory] theorists must have some underlying causal story in mind about why the patterns (again, not just any patterns, but WARP, SARP, or GARP-pruned patterns) that were found in the original choice data could credibly, or reliably, or justifiably, be extended to other data sets – why the patterns in the consumer's future choices should exhibit the same structural relationships as those found in (the pruned) past choices – but there really is no such story (explicit or implicit). CRPT is not just mindless economics (economics that is not concerned with the neurological processes that cause choice behavior): it is causeless economics.

Recall that, per the characterization of Section 2, contemporary revealed preference approaches do not represent the causal mechanisms that produce choice behavior. Thus, the theorist does not have evidence – evidence which would take the form of causes or causal mechanisms – of the stability of the relevant preferences which she can point to in order to justify her assumption of that stability. To do so would be to violate the relevant interpretation of the revealed preference relation. Of course, inductive worries can be raised about such causal mechanisms. Still, even the inductive skeptic should admit that different low degrees of confidence that the future will resemble the past are warranted by different amounts of information about the past. The identification of a single causal mechanism across different circumstances of choice behavior should increase one's confidence that the future will resemble the past, compared with the mere observance of a pattern in choice behavior across such circumstances.

To avoid the inductive skeptical puzzle, the theorist could take the domain of the revealed preference relation to be past, future, and perhaps possible choice behavior. On this solution, the revealed preference theorist who gloms on to the correct revealed preference relation is justified in

using it to predict future choices, since the revealed preference relation already represents the same choice that it aims to predict. Note that, for this solution to work, the revealed preference relation must contain the choice to be explained. Otherwise, the inductive puzzle can be raised again: what justifies the prediction of my coffee choice this morning on the basis of past coffee choices? However, if the revealed preference relation already represents the choice to be predicted, then the deep inductive puzzle set out by Hands above is, in a sense, stipulated away. "Look," the revealed preference theorist might say to the skeptic, "if you grant that I've inferred the correct revealed preference relation from an agent's past choices, then you have to grant that I am no less justified in using it to predict the future than a scientist who uses a psychological interpretation of preference, as the revealed preference relation already represents that future choice."[13]

A second solution, by contrast, maintains that the domain of the revealed preference relation is a set of (observed) actual choice behavior. It then adds the additional assumption that the agent has a stable revealed preference, i.e., that the pattern of choice does not change in future, assuming that the agent is presented with the same options.[14] In the previous case of my morning cappuccino, this solution would assume that my beverage of choice does not change as the time of day changes, and predict that I order a cappuccino in the afternoon when faced with the same options.

This assumption does prevent the letter of the circularity objection, since the revealed preference relation does not contain the choice itself.[15] And yet, the distinction between the two

---

[13] Again, the problem of the conditions under which the revealed preference theorist is justified in asserting that she has inferred future facts on the basis of past observation, given the threat of inductive skepticism, is a hard and trenchant one. But it is a hard and trenchant problem for everyone.

[14] As Binmore (2009: 9) says, "In making such inferences [about what an agent will buy in the future], two assumptions are implicitly understood. The first is that Pandora's choice behavior is stable."

[15] While it may look *ad hoc*, it's worth noting that the second solution is in keeping with the development of empirical approaches in economics, particularly in relation to Afriat's Theorem (see Afriat (1967) and also Krebs (2013: Chapter 4)

solutions is a distinction without much of a difference. The package of a revealed preference relation defined over a domain of actual, past choices and the assumption of a stable pattern of choice gets around the problem of predicting future choices in a similar way to the first solution, namely, by adding the information that the choice to be explained is part of the pattern represented by the revealed preference relation. The only difference is that the relevant work is done by a background assumption, rather than the revealed preference relation itself.

To sum up: in order to avoid a particularly pressing inductive puzzle, the revealed preference relation must be understood either to contain the choice to be explained, or as paired with a background assumption of stable preferences. This avoidance comes, however, at the cost of a violation of the No Self-Explanation Principle, and an associated, seemingly vicious circularity.

4. The causal explanation objection

A second major locus of criticism centers around repealed preference approaches' purported inability to causally explain choice. The lack of causal explanation by revealed preferences seems to follow straightforwardly from the interpretation of the revealed preference relation as a mere summary of choice behavior. The revealed preference theorist, for example, does not seem able to say that Annie chose the slice of chocolate cake *because* (in the causal sense of "because") she prefers chocolate cake. To do so would be to say something like "Annie chose the slice of chocolate cake because she chose slices of chocolate cake in similar situations in the past," or, even worse,

---

for discussion). Afriat's Theorem states that sets of choices that satisfy a certain formal constraint – the Generalized Axiom of Revealed Preference (GARP) – are representable – or, more technically, rationalizable – by a binary preference relation with certain properties. The theorem itself, however, makes no claims about choices outside of the sample domain, and it is natural to supplement the results of Afriat's Theorem with an assumption of stable preferences in order to make out of sample predictions.

something like "Annie chose the slice of chocolate cake because she chose the slice of chocolate cake." Neither gloss seems to tell a compelling causal story. The claim that choices cause other choices except by way of a mental intermediary seems like a highly implausible view, and the view that an event could cause itself seems even worse. And so, since revealed preference approaches do not represent the causes of choice behavior, revealed preference frameworks cannot be used to give causal explanations.

Of course, there are cases of scientific explanations that give generic information about causes but do not represent the particular causes of some outcome of interest. Some explanations in biology are of this sort. A biologist might explain the presence of a trait in a population due to certain selective pressures, thereby ruling out that the trait is a spandrel. Explanations such as these represent causes of a certain type (in this case, selective pressures) but do not pick out particular causal mechanisms.

Such causal abstraction is not, however, what is going on in the case of explanations that appeal to revealed preferences, on the interpretation of revealed preferences given by Section 2. For the analogy with evolutionary explanations to help the revealed preference theorist, revealed preference explanations would have to be interpreted as representing types of causes of choice behavior, while not picking out any particular underlying causes or mechanisms that produce choice, such as psychological or neurological mechanisms (see Section 2).

There is, however, a better defense of the causal-explanatory power of revealed preferences, one that deserves much more exploration than this paper can undertake. There is a package of views on which the revealed preference relation seems to causally explain choice behavior. The first component of this package is an interpretation of the revealed preference relation as a behavioral disposition. On this interpretation, the revealed preference relation entails a series of counterfactuals

of the following form: if the agent were in context *K*, she would choose option *x*. The second component is a particular view of causal explanation, namely, interventionism (see, for example, Pearl (2000) and Woodward (2003)). Interventionist accounts of causation and causal explanation represent causes and effects as variables in a causal model, and take a variable *C* to be cause of another variable *E* if and only if a suitable intervention on *C* that changes its value (and the value of no other variables) produces a corresponding change in the value of *E* (or the probability distribution of *E*). For the interventionist, the counterfactuals entailed by the revealed preference relation are more perspicuously understood as claiming that an intervention on the context variable causes a change in the choice variable.

Thus, while the causal explanation objection appears devastating, there may be a package of views that saves the causal explanatory power of revealed preferences. The rest of this paper will focus on a unificationist defense of the explanatory power of revealed preference approaches. However, I will return to this interventionist defense in the conclusion.

Suppose that you are convinced by the causal explanation objection. Is this lack of causal explanation by actual revealed preference approaches a theoretical benefit, or a damning problem? Developments in economic methodology over the last forty years or so seem to support the latter position. In particular, the discipline's move away from the formal, idealized theory-building of individuals such as Samuelson and towards causal investigation, as well as the incorporation of more psychologically realistic assumptions by behavioral economists, may cause you to doubt the introduction's assertion that revealed preference approaches (with the exception of counterfactual revealed preferences) are still common in current theoretical and empirical investigations. Those methodological developments alone, however, do not cause problems for revealed preference

approaches, as the revealed preference relation could be important for purposes other than causal explanation, such as prediction.

The addition of two assumptions, however, turns this seeming theoretical benefit into a damning problem. The first assumption is that something is an explanation in virtue of providing information about dependence relations, such as causation or laws of nature, and that this account of explanation is the single, correct account employed across the sciences. The second assumption is that, all other things equal, a framework that can provide explanations is better than one that cannot.

Interestingly, vocal defenders of revealed preference approaches, such as Binmore (2009), have exclusively denied the second assumption. The denial focuses on epistemic tradeoffs in science. As Kuhn (1977: Chapter 13), among many others, has argued, there are different epistemic virtues that scientific theories may possess, and scientists often make tradeoffs among these virtues. In the case of revealed preference approaches, Binmore (2009: 20), for example, argues that revealed preference approaches give up on the virtue of causal explanation, and gain "a theory that is hard to criticize because it has little substantive content."[16] The argument from trade-offs is particularly powerful if paired with the claim that these trade-offs are inevitable, as no single scientific theory can realize all of the epistemic virtues (see van Fraassen (1989), among others). A defense such as Binmore's concedes the heart of the criticism, that revealed preference approaches cannot provide causal explanations, and defends revealed preference approaches on other grounds. The Kuhnian-inspired defender, however, has given up the game a bit too quick, here. This tradeoff among epistemic

---

[16] The full quotation runs as follows: "The price of abandoning psychology for revealed-preference theory is therefore high. We have to give up any pretension to be offering a causal explanation of Pandora's choice behavior in favor of an account that is merely a description of the choice behavior of someone who chooses consistently. Our reward is that we end up with a theory that is hard to criticize because it has little substantive content."

virtues only gets going against a certain background assumption about explanation, namely, the first assumption above. By contrast, I think that the more successful defense of revealed preferences against the no causal explanation objection rejects this first assumption, and denies that there is indeed a tradeoff for revealed preference approaches between explanation and other virtues. Let's now turn to that defense.

5. A fundamental disagreement about explanation

I want to pause and cast an eye over the dialectic thus far. We've seen claims of generalizations as mere patterns in the phenomena, rather than something more modally robust; explanatory circularities; charges that a philosophical theory results in explanatorily inert generalizations…if this reminds you of the disagreement between Humeans and anti-Humeans over laws of nature, then I whole-heartedly agree. The analogy with the debate over laws of nature reveals a principled response to both the no self-explanation objection and the causal explanation objection: unificationism about explanation.

Three notes about terminology before I proceed. First, the analogy draws on a common epistemic property of Humean laws and the revealed preference relation, namely, their systematicity. In the case of laws, this systematicity, and the Humean account of laws more generally, is usually paired with other background commitments, such as a denial of fundamental necessities. However, as Demarest (2017) and Hicks (forthcoming) discuss, the Humean view of laws is compatible with other background commitments, such as a commitment to fundamental necessities. This point is important for the plausibility of revealed preference approaches, as contemporary defenders of revealed preferences accept the existence and causal influence of mental states on choice (see Binmore (2009: 19--20) for one example). To separate the discussion here from such orthogonal

background Humean commitments (which are difficult to mentally jettison), I will follow Hicks (forthcoming) in using "regularity theorists" instead of "Humeans" and "modalists" instead of "anti-Humeans."

The second note concerns "unificationism." As Colombo and Hartmann (2017) discuss, "unification" is a heterogeneous notion in science, one which often comes apart from explanation. For present purposes, two notions of unification are relevant. The first notion of unification is that of modeling a set of diverse phenomena within a single framework or with a few mathematical equations (see Colombo and Hartmann (2017: Section 4)). The second notion of unification is the notion of a systematization of a set of facts by showing how those facts are instances of a more general pattern (see Friedman (1974) and Kitcher (1989)).

Finally, these two different notions of "unification" are the basis for two distinct roles that unification plays in science. First, scientists often prefer theories that are more unifying, i.e., a theory that can model a more set of diverse phenomena within a single framework or set of equations. Call this the *criterion of theory choice* use of "unification." Thus far, the literature on revealed preferences and unification has focused on the criterion of theory choice use of "unification."[17] By contrast, this paper argues that revealed preference theorists should adopt a unificationist theory of *explanation*,

---

[17] Ross (2005), for example, argues that revealed preference approaches can be defended on unificationist grounds. I read Ross invoking unification in this criterion of theory choice sense, arguing that economists should aspire "to be contributors to a unified behavioral science" (2005: 219). In that vein, he develops an account of agency that can both underlie microeconomic practice and is respectable from the perspective of cognitive science. Mäki (1990, 2001, 2002) has also done extensive work on unification in economics. I read Mäki's (1990, 2001) notion of "ontological unification" in the criterion of theory choice sense of unification. It is clear that Mäki does not defend a unificationist theory of explanation as I describe in this section, as he describes ontological unification as discovering that diverse phenomena have the same "causes, origins, constituents" (2001: 498). See Lehtinen (2012: Introduction, Section 5) for discussion of Mäki on unification in economics.

where a set of information counts as an explanation in virtue of showing how the fact to be explained is an instance of a more general pattern.

5.1 The analogy with explanatory debates about laws of nature

Regularity theorists and modalists have long been embroiled in a debate over whether laws that are mere regularities explain. Here I agree with Miller (2015) that the disagreement is in part motivated by a background explanatory disagreement, and that the regularity theorist can avail herself of a unificationist theory of explanation to avoid certain explanatory criticisms that go back at least to Armstrong (1983). In Section 5.2, I will use lessons from this debate to argue that revealed preference theorists should also adopt a unificationist conception of explanation in order to avoid structurally similar criticisms.

The starting point of regularity views is the commitment that laws are mere patterns in the phenomena, rather than anything more modally robust. The regularity view that I will focus on is the Mill-Ramsey-Lewis Best System Account of laws (see Lewis (1999a: 41--43; 1999b: Section 3), Mill (1843), and Ramsey (1928)). The Best System Account posits that laws are the members of a small set of axioms that allow for the derivation of a large set of facts. For present purposes, the details of the view are not important (but for contemporary amendments see Beebee (2000), Loewer (2007), and Callendar and Cohen (2009, 2010)). Rather, we are interested in the epistemic motivation behind the Best System Account (often left implicit among regularity theorists, although explicitly discussed by Hicks (forthcoming)). For the regularity theorist, laws systematize the phenomena in question. The motivation for such systematicity is an epistemic one: one of the epistemic roles of laws is to efficiently organize the blooming, buzzing confusion and thereby act as simple tools for the

prediction and explanation of a much larger set of facts (see Hicks (forthcoming) for discussion, and for a proposal about other epistemic roles of laws).

Modalists, by contrast, take laws to be more modally robust (see Armstrong (1983), Dretske (1977), Lange (2007), and Maudlin (2007), among others). For modalists, laws constrain the dynamic evolution of physical states, as well as the state of the world at some time. In other words, the laws make the world evolve as it does, rather than merely reflecting its evolution, as the regularity theorist thinks. The modalist background conception of explanation is motivated by this view of laws of nature. Scientists explain phenomena using laws and robust generalizations. So, the modalist is naturally led to a view where explanations showcase constraints, or that in virtue of which the explanatory target had to obtain, given some information about (and perhaps further constraints on) the metaphysical structure of reality. I will call these *constraint-based theories*.

Modalists have long objected to regularity theories on the grounds that laws that are constituted by mere regularities cannot explain particular matters of fact. Here is Armstrong (1983: 40) weighing in against regularity theories (see also Maudlin (2007: 72)):

> Could this hypothesis [the law that All F's are G's] serve as an explanation [of the fact that all observed F's are G's]? It does not seem that it could. That all F's are G's is a complex state of affairs which is in part *constituted* by the fact that all observed F's are G's…As a result, trying to explain why all observed F's are G's by postulating that all F's are G's is a case of trying to explain something by appealing to a state of affairs part of which is the thing to be explained. But a fact cannot be used to explain itself.

For Armstrong, laws, as understood by the regularity theorist, are constituted by all the individual instances of that law. Since laws explain their instances, those instances partially explain themselves. (Note that this explanatory criticism mirrors the no self-explanation objection of Section 3, where revealed preference approaches were charged with invoking an instance of choice behavior to partially explain that instance.)

Following Miller (2015), I think that the regularity theorist should adopt a background unificationist view of explanation to escape the above explanatory criticism (see Friedman (1974) and Kitcher (1989) for example unificationist theories of explanation). On the unificationist view, "explanation is sometimes a matter of uncovering mere patterns or regularities in reality, and then classifying particular facts as instances of these patterns" (Miller 2015: 1326). In other words, explanations systematize a set of facts by showing how some of those facts are instances of a more general pattern.

Constraint-based theories and unificationist theories of explanation are not distinguished merely by the type of information given by explanations.[18] Indeed, the unificationist theory can, and should, say that explanations sometimes give information about laws, causes, and other dependence relations. Instead, the two theories disagree about *why* a set of information is explanatory. To bring out the difference, let's consider the explanatory question of why a particular sample of sodium chloride burns yellow. For both the unificationist regularity theorist and the constraint-based modalist, the generalization "all sodium chloride samples burn yellow in a certain range of background conditions" partly answers the explanatory question. According to unificationist theories of explanation, "all sodium chloride samples burn yellow in a certain range of background conditions" is explanatory because it unifies a set of instances by picking out a macrostructure – namely, being a sample of sodium and chloride atoms bound in a particular configuration – that they share. In other words, one way of unifying (in a distinctly explanatory fashion) is by subsuming an

---

[18] Of course, the two theories will also sometimes differ about which generalizations are explanatory. For the unificationist, some universal generalizations that describe a property shared by tokens of some type – such as "all hydrogen atoms have mass *m*" (see Miller (2015: 1325)) – will count as explanatory, even if they are not shorthand for information about causes or laws that metaphysically determine that property. For constraint-based theorists, such generalizations must be understood as shorthand for information about the causes or laws that metaphysically determine that property.

instance under a pattern by identifying structural features of the instance that it shares with other instances (see Miller (2015: 1325)). According to constraint-based theories of explanation, by contrast, "all sodium chloride samples burn yellow in a certain range of background conditions" is explanatory because it provides information about the features of the world that determine this behavior, namely, the macrostructure of sodium chloride, which causes it to burn yellow.

Unification has a clear explanatory importance, one reflected in current and past scientific practice.[19] One such example is Johannes Kepler's and Isaac Newton's respective explanations of the motion of the planets in our solar system. Between 1609 and 1619, Kepler published three laws of planetary motion that, building on work by Nicolas Copernicus and others, were significant advances on the previous Ptolemeic tradition. For example, Kepler proposed that the orbit of the planets around the Sun are ellipses, with the Sun at one focus. Kepler's three laws, however, are tailored to the motion of the planets. In 1687, Newtonian demonstrated that Kepler's three laws are, to a good approximation, a consequence of his laws of motion and law of universal gravitation. Newton's laws are unifying in the first sense of "unification" discussed previously: they apply to all objects in the universe, i.e., a more diverse set of phenomena than Kepler's laws. They are also unifying in the second sense of "unification" discussed earlier: they subsume planetary motion under a widespread pattern. They are also a widely agreed upon explanatory advance over Kepler's laws.[20]

The above example motivates an important connection between unification and explanation, but does not yet motivate a unificationist theory of explanation. Indeed, proponents of constraint-

---

[19] For an argument that unification is not explanatorily significant, see Morrison (2000).

[20] Kepler's laws are sometimes described as descriptions of planetary motion, rather than explanations thereof. The characterization of Kepler's laws as descriptions or explanations works equally well for the arguments here, since the key point is that Newton's laws, which are more unifying, are explanatorily superior to Kepler's laws.

based theories can – and have (see Strevens (2008), for example) – incorporate the explanatory importance of unification into their theories. The reason that I will not motivate and commit to a particular unificationist theory is two-fold. First, the explanatory defense of revealed preference approaches requires a view of explanation with the following two properties: (1) an account on which something is an explanation in virtue of epistemic facts, rather than facts about explanatorily relevant metaphysical constraints, such as causes, and (2) some of those epistemic facts are facts about some type of systematicity. Those two properties underdetermine a particular view of explanation.[21] Second, it must be admitted that the most prominent unificationist account (Kitcher (1989)) has been devastatingly criticized, particularly for its inability to account for explanatory asymmetries (see Barnes (1992) and Woodward (2003), among others). A full defense of revealed preference approaches along the lines recommended in Section 5.2 requires the development of a unificationist theory of explanation, which I will not undertake here.

5.2 Unificationism to the rescue

Like the regularity theorist about laws, the proponent of revealed preference approaches should adopt a unificationist account of explanation. Unificationism about explanation is a better fit with many of the motivations and background commitments of revealed preference approaches than a constraint-based view. More crucial for the current project, however, is that this move allows the

---

[21] In other words, the unificationist about explanation need not maintain that unification is a matter of the derivability of a large set of phenomena from a small set of axioms (i.e., need not be a "derivational unificationist," as Mäki (1990, 2001) puts it).

proponent of revealed preferences to escape the explanatory criticisms of Sections 3 and 4.[22] This section argues that last point.

The response to the criticisms of Sections 3 and 4 comes in two parts: a denial of the No Self-Explanation Principle, and a denial of the claim that explanations must give information about dependence relations. I will begin with the latter claim, since its denial is entailed by unificationist accounts of explanation. As discussed in Section 5.1, unificationist and constraint-based views of explanation fundamentally disagree about why a set of information is an explanation. For the unificationist, a set of information is explanatory if and only if and because it is appropriately unifying; for constraint-based views, a set of information is explanatory if and only if and because it contains information about the explanatorily relevant metaphysical constraints on the target phenomenon. While the unificationist can allow for explanations that give information about dependence relations, she will maintain that explanations need not do so. That is because, according to her theory, facts about explanation supervene on facts about unification, rather than on facts about dependence relations.

Unificationism pairs nicely with the epistemic purposes and representational goals that motivated the introduction of the revealed preference relation into choice-based frameworks. Recall the discussion in Section 2 of one epistemic goal behind the introduction of the revealed preference relation: to capture patterns among choice behavior in an efficient manner. Both unificationism and the revealed preference methodology focus on systematicity as an epistemic virtue. Furthermore, as was argued in Section 4, revealed preferences approaches may not give causal explanations. Indeed,

---

[22] Along with this normative claim, I also think that many revealed preference theorists have been working with a background, unarticulated unificationist concept of explanation, but I will not defend the psychological claim here.

recall that Binmore (2009: 19--20) argues not only that revealed preference frameworks do not give causal explanations, but also that it is an advantage of such frameworks that they do not. Revealed preference approaches, then, are a much better fit with a unificationist view of explanation than with a constraint-based view.

Let's now turn to the issue of self-explanation. What is the relationship between unificationism and the No Self-Explanation Principle, and how might a unificationist motivate the denial of that principle? Unificationism does not entail the permissibility of self-explanation. However, as Miller (2015) has argued, unificationist explanation is friendly to self-explanation.

Consider first an example from Miller (2015: 1325). We might explain why a hydrogen atom has a certain mass m by appeal to the fact that all hydrogen atoms have mass m. Arguably, the universal fact that all hydrogen atoms have mass m is, in turn, explained by all the instances of hydrogen atoms with mass m. So, by transitivity, we have a case of self-explanation: the fact that this hydrogen atom has mass m partially explains why it has mass m. From a constraint-based perspective, this certainly looks like no explanation at all, since we have not learned why all hydrogen atoms must have mass m. If there are no constraints on facts about the mass of hydrogen atoms, then they have no explanation – they are brute facts.

From a unificationist perspective, however, self-explanation appears much less fraught. Once you adopt the unificationist mindset, it is no longer problematic to explain the fact that this hydrogen atom has mass m by noting that it has mass m; another hydrogen atom over there has mass m; a third has mass m; and so on and so forth. There are no metaphysical grounds to exclude self-explanation, since explanations do not have to give information about irreflexive constraints. There are no epistemic grounds either. Explanations, after all, are efficient summaries of patterns among particular matters of fact. Sometimes, we gain more efficient summaries by formulating

generalizations that include the particular fact itself, as in the example of why a hydrogen atom has a particular mass *m* above. Or, indeed, in the case of the revealed preference relation: it is by fitting a particular instance into a pattern of choices that the revealed preference theorist gets around the inductive puzzle of Section 3 to predict future choices, for example. Therefore, the unificationist generally, and the revealed preference unificationist in particular, should not adopt the No Self-Explanation Principle, since her theory relies on the explanatory strategy of fitting instances into a pattern of which those instances are members, i.e., self-explanation.

As the conclusion will elaborate upon, there remains a great deal of further work to be done if a positive explanatory case is to be made for revealed preference approaches. This section took the first step of that positive case: to defend revealed preference approaches against the explanatory objections of Sections 3 and 4. In so doing, it also suggested that unificationism about explanation and revealed preference approaches are a promising explanatory combination. That topic will now be taken up in the conclusion.

6. Where to now?

This paper has argued for the following conditional conclusion: if the revealed preference theorist adopts a background unificationist theory of explanation, then she avoids the two explanatory problems posed in Sections 3 and 4. The pairing of revealed preference approaches with unificationism shows that it is wrong to consider the revealed preference modeler as tasked with the construction of models that capture underlying (causal) constraints. Instead, the revealed preference modeler attempts to find an efficient systematization of a set of choices. In other words, she looks for mere patterns among choice behavior. By her own lights, a proponent of revealed preference

approaches succeeds by doing so, even if she does not provide a causal explanation of choice, and she violates the No Self-Explanation Principle.

I'll end by discussing two natural next steps in the dialectic. An important question that arises out of this paper is whether revealed preference approaches are, indeed, more unifying. That question, of course, requires a comparison – more unifying than what alternatives? In the sense of efficient expression, a choice-based framework with a revealed preference relation is clearly more unifying than a choice-based framework without one. The harder question for the proponent of revealed preferences is whether revealed preference approaches are sometimes more unifying than frameworks that represent psychological facts. Proponents of revealed preference approaches, such as Binmore, certainly take them to be so. Here is Binmore (2009: 8--9):

> Modern decision theory succeeds in accommodating the infinite variety of the human race within a single theory simply by denying itself the luxury of speculating about what is going on inside someone's head. Instead, it pays attention only to what people do. It assumes that we already know what people choose in some situations, and uses this data to deduce what they will choose in other situations.

Note that the defense of revealed preference approaches cannot merely claim that *bracketing* causes at the individual level is more unifying, for reasons related to the discussion of evolutionary explanations in Section 4. There, I pointed out that many causal explanations do not explicitly represent causes or causal-mechanisms at the individual level. Such explanations look to be more unifying in virtue of doing so. However, the biologist would not deny that these explanations stand in for specific causal mechanisms in particular populations. By analogy, a mentalist about preferences can maintain that economic explanations are better in virtue of bracketing causes, but that these explanations represent mental entities which cause choice behavior. The revealed preference theorist thus has to argue that her framework is more unifying in virtue of refusing to represent mental preferences at any stage in the modeling process.

Binmore has articulated a compelling thought here, but it is one that requires more defense. There are two complimentary routes to pursuing such a defense, one more epistemic and the other more metaphysical. The more epistemic defense argues that, due to cognitive limitations of human beings, the diversity of their cognitive interests, and facts about their representational and technological capabilities, unificationist explanations that do not aim to represent the psychological determinants of choice allow us to better explain, and thereby understand, large-scale explanatory patterns that would be less readily apparent at the psychological level. This defense can look defeatist: in some cases, our cognitive and technological capacities are too limited to get at the causal structure of the world, so we seek efficient summaries of patterns instead. However, one might lean more heavily on the diversity in our cognitive interests, rather than the limitation in our capabilities, in order to present a less defeatist version of this defense.[23] Furthermore, the epistemic defense appears less defeatist when paired with a metaphysical defense that starts from the fact that some systems in the world, such as the large social systems studied by economists, are very complex. This complexity sometimes grounds the explanatory superiority of unificationist over constraint-based explanations, as simple pattern representation allows inquirers to abstract away from much of the complexity.

Of course, these two routes are nicely complimentary. And indeed, together, they offer an argument for explanatory pluralism, one starting from a case study of revealed preference approaches. Such a two-pronged strategy fits nicely, for example, with arguments from Kellert, Longino, and Waters (2006: Introduction) that such case studies "establish the possibility that the world is too complicated or too indeterminate and our cognitive interests too diverse for the monist

---

[23] Kellert, Longino, and Waters (2006: xi) discuss the diversity in our cognitive interests as a motivation for pluralism.

ideals." However, as they stand, the arguments of the paper do not yet support explanatory pluralism. A successful argument for explanatory pluralism requires an unconditional defense of the pairing of unificationism and revealed preference approaches. Furthermore, the question of whether interventionism about explanation equally well or better accounts for the explanatory practice of revealed preference approaches is still an important open question, raised but not settled here. If, indeed, it does, such a fact speaks against revealed preference approaches as a case study that supports explanatory pluralism.

This last point brings us back to a very different avenue of explanatory defense for revealed preference approaches: interventionism. Section 4 suggested that hypothetical revealed preferences seem to be causes, at least against a background interventionist view of causation. If so, the combination of the interpretation of revealed preferences as a behavioral disposition and an interventionist view of causal explanation offers a potential second defense of the explanatory power of revealed preference approaches. This defense has two main advantages. It avoids theoretical challenges, such as the problem of causal explanation. It also reconciles revealed preference approaches with economics' current methodological focus on causal modeling. The next step in the dialectic is to address the following two questions. First, if interventionist views of causation do deem revealed preferences to be causes, is that an objection to interventionist views, or a plausible result? Second, even if revealed preferences are causes, are such putative revealed preference explanations (good) explanations? But, I'm afraid that such questions will have to wait for another time.

References

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

Agner, Erik. forthcoming. What preferences really are. *Philosophy of Science.*

Armstrong, D. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

Barnes, E. 1992. Explanatory unification and the problem of asymmetry. *Philosophy of Science* 59(4): 558--571.

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

Beebee, H. 2000. The non-governing conception of the laws of nature. *Philosophy and Phenomenological Research* 61: 571--594.

Bernheim, B. and A. Rangel. 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics* 124(1): 51--104.

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

Binmore, K. 2009. *Rational Decisions.* Princeton: Princeton University Press.

Blundell, R. 2005. How revealing is revealed preference? *Journal of the European Economic Association* 3: 211--235.

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

Callender, C. and J. Cohen. 2009. A better best system account of lawhood. *Philosophical Studies* 145(1): 1--34.

Callender, C. and J. Cohen. 2010. Special science, conspiracy, and the better best system account of lawhood. *Erkenntnis* 73: 427--447.

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

Colombo, M. and S. Hartmann. 2017. Bayesian cognitive science, unification, and explanation. *British Journal for the Philosophy of Science* 68(2): 451--484.

Demarest, H. 2017. Powerful properties, powerless laws. In *Putting Powers to Work: Causal Powers in Contemporary Metaphysics*, ed. J. Jacobs. Oxford: Oxford University Press.

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.

Dretske, F. 1977. What is a law of nature? *Philosophy of Science* 44: 248--268.

Gul, F. and W. Pesendorfer. 2008. The case for mindless economics. In *The Foundations of Positive and*

Normative Economics, eds. A. Caplin and A. Shotter. Oxford: Oxford University Press.

Friedman, M. 1974. Explanation and scientific understanding. *Journal of Philosophy* 71(1): 5--19.

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

Guala, F. 2012. Are preferences for real? Choice theory, folk psychology, and the hard case for commonsense realism. In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, ed. A. Lehtinen, J. Kuorikoski, and P. Ylikoski, Part II, Chapter 6. Abingdon: Routledge.

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

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

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

Hempel, C. and P. Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science* 15: 135--175.

Hicks, M. forthcoming. Dynamic Humeanism. *The British Journal for the Philosophy of Science*.

Hildenbrand, W. 1994. *Market Demand*. Princeton: Princeton University Press.

Kellert, S., H. Longino, and C. K. Waters. 2006. Introduction. In *Scientific Pluralism*, eds. Kellert, S., H. Longino, and C. K. Waters, vii--xxix. Minneapolis: University of Minnesota Press.

Kitcher, P. 1989. Explanatory unification and the causal structure of the world. In *Scientific Explanation*, eds. P. Kitcher and W. Salmon, 410--505. Minneapolis: University of Minnesota Press.

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

Kuhn, T. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension*, 320--339. Chicago: The University of Chicago Press.

Lange, M. 2009. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford: Oxford University Press.

Lehtinen, A. 2012. Introduction. In *Economics for Real: Uskali Mäki and the Place of Truth in Economics*, ed. A. Lehtinen, J. Kuorikoski, and P. Ylikoski, Introduction. Abingdon: Routledge.

Lewis, D. 1994a. New work for a theory of universals. *Papers in Metaphysics and Epistemology*, 8--55. Cambridge: Cambridge University Press.

Lewis, D. 1994b. Humean supervenience debugged. *Papers in Metaphysics and Epistemology*, 224--247. Cambridge: Cambridge University Press.

Loewer, B. 2007. Laws and natural properties. *Philosophical Topics* 35(1/2): 313--328.

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.

Mäki, U. 2002. Explanatory ecumenism and economic imperialism. *Economics and Philosophy* 18: 237-259.

Mas-Collel, A., M. Whinston, and J. Green. 1995. *Microeconomic Theory*. Oxford: Oxford University Press.

Maudlin, T. 2007. *The Metaphysics Within Physics*. Oxford: Oxford University Press.

Mill, J.S. 2009 [1843]. *A System of Logic, Ratiocinative and Inductive*. Reprint. New York: Echo Library.

Miller, E. 2015. Humean scientific understanding. *Philosophical Studies* 172(5): 1311--1332.

Morrison, M. 2000. *Unifying Scientific Theories: Physical Concepts and Mathematical Structure*. Cambridge: Cambridge University Press.

Pearl, J. 2000. *Causality*. Cambridge: Cambridge University Press.

Ramsey, F. W. 1990 [1927]. Laws and causality. Reprinted in *F. P. Ramsey: Philosophical Papers*, ed. D. H. Mellor, 140--164. Cambridge: Cambridge University Press.

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

Ross, D. 2014. *Philosophy of Economics*. London: Palgrave Macmillan.

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

Scriven, M. 1962. Explanations, predictions and laws. *Minnesota Studies in the Philosophy of Science* 2: 170--230.

Sen, A. 1977. Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs* 6(4): 317--344.

Strevens, M. 2008. *Depth*. Harvard: Harvard University Press.

van Fraassen, B. 1989. *Laws and Symmetry*. New York: Oxford University Press.

Varian, H. 1982. Nonparametric test of models of consumer behavior. *Review of Economic Studies* 50: 99--110.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.