

Preference as Desire

James Fanciullo

Lingnan University

Forthcoming in the *Journal of Philosophy*

Abstract: This paper considers two competing views of the relationship between preference and desire. On what I call the “preference-first view,” preference is our most basic form of conative attitude, and desire reduces to preference. This view is widely assumed, and essentially treated as orthodoxy, among standard decision theorists, economists, and others. I argue, however, that the preference-first view has things the wrong way around. I first show that the standard motivation offered for this view—motivation underlying foundational work in decision theory and economics—leaves the view with unacceptable psychological implications. I then introduce an alternative view—the “desire-first view”—on which desire is our most basic form of conative attitude, and preference reduces to desire. On the desire-first view I propose, preferences, as comparisons, are best understood as comparisons of the extents to which alternatives are desired. I show that this desire-first view is simple, ecumenical, and explanatorily powerful.

It is widely assumed that human decision making is largely a matter of certain cognitive and conative mental states properly combining. This very general assumption underlies a great deal of work in philosophy, cognitive science, economics, psychology, decision theory, and elsewhere.¹ Cognitive

¹ See, for example, Daniel M. Hausman, *Preference, Value, Choice, and Welfare* (Cambridge: Cambridge University Press, 2012); John L. Pollock, *Thinking About Acting: Logical Foundations for Rational Decision Making* (New York: Oxford University Press, 2006); Timothy Schroeder, *Three Faces of Desire* (New York: Oxford University Press, 2004); Armin W. Schulz, “Preferences vs. Desires: Debating the Fundamental Structure of Conative States,” *Economics and Philosophy*, XXXI, 2 (July 2015): 239-57; and Katie Steele and H. Orri Stefánsson, “Decision Theory,” in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), <https://plato.stanford.edu/archives/win2020/entries/decision-theory/>, among many others.

states, functioning to represent the world, combine with conative states, which function to represent how the world is to be, creating a decision as to how to act.

The standard approach in decision theory and economics especially refines this picture. In light of foundational work in these fields, it is standardly believed, or at least assumed, that the fundamental conative facts in this picture are facts about an agent's *preferences*.² To model rational decision making, on this picture, we must first model an agent's preferences between different alternatives, as these preferences represent the agent's fundamental conative or evaluative states. Only then can we determine further facts about things like what the agent desires, the extent to which they desire different things, and ultimately, in combination with the agent's beliefs, what it is rational for the agent to do. On the standard approach in decision theory and economics, then, preference is our most basic form of conative attitude, and desire is to be explained in terms of preference. Call this the *preference-first view*.

The decision theorist's motivation for adopting this view is roughly as follows.³ On the one hand, decision-theoretic reasoning requires a cardinal measure of value: the extent to which an option is valued by an agent is a fundamental part of calculating an option's expected utility. A mere ordinal measure of value—or a ranking with no specification of the extent to which one item ranks above another—would not be enough. Here, desires might seem to play a natural role: as anyone with desires knows, we can desire some things very much, and others very little. On the other hand, however, decision-theoretic reasoning requires that this cardinal measure be in some way introspectable, and we don't seem capable of introspecting a cardinal measure of our desires. We can perhaps introspect that we desire some things very much and others very little, but not that we

² For overviews here, see Pollock, *Thinking about Acting*, *op. cit.*; and Steele and Stefánsson, "Decision Theory," *op. cit.*

³ For overviews here, see Pollock, *Thinking about Acting*, *op. cit.*; Timothy Schroeder, "Desire and Pleasure in John Pollock's *Thinking about Acting*," *Philosophical Studies*, CXLVIII, 3 (April 2010): 447-54; and Timothy Schroeder, "Desire," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), <https://plato.stanford.edu/archives/sum2020/entries/desire/>.

desire a given thing to some specific, numerical degree (like 9, or 103). In contrast, we do seem capable of introspecting our simple ordinal ranking of any two options: we know which we would take between the two. Here, preferences might seem to play the natural role: we can introspect facts about what we *prefer*. The decision theorist's problem, then, is that while decision-theoretic reasoning requires a cardinal measure of value, this cardinal measure must ultimately be based on something introspectable. Thus, their problem would be solved if they could derive the cardinal measure from the introspectable ordinal measure. And this is just what many have done. Specifically, a great deal of influential work in decision theory has focused on proving that, if an agent's total set of preferences satisfies various constraints and axioms, it is possible to derive facts about the extent to which any given option is valued by the agent.⁴ The basic conative facts in decision theory, on this standard picture, are thus facts about an agent's preferences. Underlying all this work, then, is the preference-first view.

These proofs—or “representation theorems”—are fascinating, and extremely theoretically important. However, of course, they go little way toward showing what human evaluative psychology is actually like. As we've just seen, the preference-first view is adopted not in virtue of its independent plausibility, but instead in virtue of certain constraints on the decision theorist's ultimate project. In particular, the decision theorist happens to need to derive a cardinal measure from something introspectable, such as a preference ranking. But, as John Pollock nicely puts the point:

Regardless of what else they may establish, it must be emphasized that representation theorems show nothing directly about how human beings perform practical cognition. In particular, they clearly do

⁴ See, perhaps most famously, Frank P. Ramsey, “Truth and Probability,” in R. B. Braithwaite, ed., *The Foundations of Mathematics and other Logical Essays* (London: Kegan, Paul, Trench, Trubner & Co., 1926), pp. 156-98; Leonard J. Savage, *The Foundations of Statistics* (New York: John Wiley, 1954); and John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (New York: Wiley, 1944).

not show that humans proceed by recovering a cardinal measure from their preference rankings and then use it to reason decision-theoretically. Despite this, I have found that a huge number of philosophers, economists, decision analysts, psychologists, and other cognitive scientists believe that this is an accurate account of human evaluative cognition. Specifically, they believe that what is basic to evaluative cognition is a set of binary preferences, and all other aspects of evaluative cognition must derive from these preferences.⁵

Just as the theorems prove little about “human evaluative cognition,” so they prove little about the nature of our basic form of conative attitude, or indeed about the relationship between preference and desire. So, we can ask: is the human mind truly this way? Is preference our most basic form of conative attitude? Or is this dominant view in fact mistaken?

Surprisingly, very little work has been done with the aim of answering these questions, or of analyzing the independent plausibility of the preference-first view or its potential rivals.^{6,7} My aim in this paper is to help fill this lacuna. I’ll do this, first, by briefly analyzing the decision theorist’s above motivation for the preference-first view, and pointing out two assumptions about preference and desire that play crucial roles in this line of reasoning. I’ll then go on, in section 2, to discuss one of the very few explicit arguments against the preference-first view of which I’m aware, due to John Pollock, as well as a recent rebuttal of his argument. As I’ll show, while Pollock’s argument fails to disprove the preference-first view itself, it succeeds in undermining the decision theorist’s primary motivation for the preference-first view. We’ll thus be left with no reason to reject the preference-first view, but also no reason to accept it. Then, in sections 3 and 4, I’ll tip the scales. In particular,

⁵ Pollock, *Thinking about Acting*, *op. cit.*, p. 23.

⁶ The few exceptions here include José Luis Bermúdez, *Decision Theory and Rationality* (Oxford: Oxford University Press, 2009); Pollock, *Thinking about Acting*, *op. cit.*; and Schulz, “Preferences vs. Desires,” *op. cit.* See also Schroeder, “Desire and Pleasure,” *op. cit.* and Schroeder, “Desire,” *op. cit.* for discussion of Pollock’s arguments especially.

⁷ Interestingly, there is a lively, parallel debate in epistemology regarding the relationship between credence and belief—see, for example, Elizabeth G. Jackson, “The Relationship Between Belief and Credence,” *Philosophy Compass*, XV, 6 (June 2020): e12668.

I'll introduce, motivate, and defend a rival view, which I'll call the *desire-first view*. As the name indicates, this view claims that desire is our most basic form of conative attitude, and preference is to be explained in terms of desire. I'll develop a version of this view, illustrate its advantages, and conclude that the standard view of the basic conative structure of human minds should be rethought.

1. Why accept the preference-first view?

The preference-first view, as I say, has pride of place in economics, decision theory, and elsewhere. As we've seen, this in large part due to certain constraints on the decision theorist's ultimate project. In outlining this general state of play, however, I've glossed over two important, widely held assumptions about the natures of preference and desire that have helped lead the decision theorist to the preference-first view. These are two presumed differences between preference and desire (despite their many apparent similarities, which we'll discuss below).⁸

The decision theorist, recall, requires an introspectable, cardinal measure of value. In preference, we've seen, they take themselves to have something perfectly introspectable. This is because, for any two alternatives, it seems we can easily introspect which of the two we prefer. Notice, though, that this reveals something about preferences: that they are essentially *comparative*.⁹ To prefer is always to prefer one thing to another. To simply say "I prefer A" is akin to saying "the candidate is better" or "the noise is louder." Outside of context, these claims make little sense, and are at best incomplete. This is because they're failed attempts at making comparisons. Preference, in

⁸ For the standard decision theorist and economist, perhaps the most notable similarity here is that (as conative attitudes) both preference and desire seem to entail motivation, or dispositions to act so as to make their objects more likely to obtain.

⁹ See, for example, John Broome, *Ethics out of Economics* (Cambridge: Cambridge University Press, 1999); Daniel M. Hausman, "Mistakes about Preferences in the Social Sciences," *Philosophy of the Social Sciences*, XLI, 1 (March 2011): 3-25; Hausman, *Preference, Value, Choice, and Welfare*, *op. cit.*; and Schulz, "Preferences vs. Desires," *op. cit.* In fact, I am unaware of anyone who thinks (or argues) otherwise.

this way, is unlike desire: to say “I desire A” is to say something complete, and requires no further context. Thus desires, we might say, are essentially *monistic*: they have a single intentional object. In contrast, preferences are essentially *binary* or (as I’ll say) *comparative*: they have two intentional objects. While both are evaluative attitudes, a crucial introspectable feature of preference, for the decision theorist, is that preferences compare, and in particular rank, alternatives. After all, as long as these rankings are available, the decision theorist has all the basic evaluative information they need to model decision-making—nothing more complex needs to be introspected.

This brings us to the second apparent difference between preference and desire. The decision theorist, as I say, can get all the evaluative information they need from an introspectable preference ranking. This is because, while they ultimately need a cardinal measure of value—or a measure of the extent to which given items are valued—their “representation theorems” show that they can derive this cardinal measure from (what they take to be) the more basic, introspectable ordinal measure. Hence the second presumed feature of preference that separates it from desire, on the standard decision theorist’s picture, is that preferences provide merely *ordinal* rankings. Indeed, this is the crucial introspectable feature of preferences: that they rank—and in particular ordinally rank—one thing above another. I can easily introspect, for instance, both that I prefer my sister’s being healthy to my eating a piece of chocolate, and that I desire both of these things. While I cannot introspect a precise cardinal measure of my desires, though, notice that my preference alone provides a precise, introspectable ordinal ranking. The decision theorist’s constraints regarding introspectability thus lead them to take as basic the introspectable, ordinal rankings provided by preferences, from which they can go on to derive—via various representation theorems—all further evaluative facts relevant to decision-making, including facts about the cardinal measure of any given desire. The second presumed difference between preference and desire, then, is that whereas desires fundamentally provide a (non-introspectable) cardinal measure of value, or are fundamentally *graded*,

preferences fundamentally provide an (introspectable, merely) ordinal measure of value, or are fundamentally *non-graded*.¹⁰

These represent the standard decision theorist's two presumed differences between preference and desire: that preferences are essentially comparative, whereas desires are essentially monistic; and that preferences are fundamentally non-graded, whereas desires are fundamentally graded. As we've seen, the decision theorist's constraints regarding introspectability play a large role in their acceptance of these differences. Moreover, these differences play a crucial role in the decision theorist's ultimate project. The hugely influential representation theorems offered here purport to show precisely that, given just facts about our binary, non-graded conative attitudes, we can derive all facts about our monistic, graded conative attitudes. Without these assumptions about preference—as well as the constraints on introspectability—there would have been no motivation for the representation theorems in the first place. This view of preference thus leads the decision theorist to the representation theorems, and thus leads them to what I've called the preference-first view.

Again, however, the fact that this view of preference allows the decision theorist to proceed with her project tells us little about whether this view accurately represents the conative structure of our minds. While surprisingly little attention has been paid to this fact, at least one theorist, John Pollock, has explicitly argued that this conative structure cannot be as the decision theorist describes it, and hence that the preference-first view is false. In the following section, I'll explore Pollock's argument, as well as a recent rebuttal of it, and determine what it ultimately teaches us about the preference-first view.

¹⁰ I borrow the “graded” terminology from Schulz, “Preferences vs. Desires,” *op. cit.*

2. Pollock and Schulz on the preference-first view

Pollock's argument against the preference-first view is an appeal to computational tractability.¹¹ The standard decision theorist takes preference to be an essentially non-graded, comparative attitude, and desire to be an essentially graded, monistic attitude. Their representation theorems show that it is possible to derive all facts about the graded attitudes from facts about the non-graded attitudes. According to Pollock, however, even if the theorems prove that this is possible in theory, it is a wholly implausible picture of how the evaluative facts are realized in human brains. This is because the number of basic preferences (so understood) that would be required to derive facts about desires (so understood) would be implausibly large. Specifically, Pollock argues, even if there were just 300 types of states of affairs that mattered to us—and, he notes, it seems likely there are many more than this—the number of basic preferences that would be needed to encode facts about the extent to which any given thing was desired would be far greater than the number of particles in the universe. Given this seems to be beyond the bounds of what the human brain can store, it seems human beings cannot be as the preference-first view seems to predict. In contrast, Pollock goes on to note, the very same facts about desires and preferences—including facts about the extent to which any given thing was desired—would require just 300 basic facts about desires. This difference, as Pollock puts it, “is the difference between the trivial and the impossible.”¹² Accordingly, he concludes, the more psychologically realistic view seems to be the one on which desire is the more basic conative attitude.

Pollock's argument here depends crucially on the standard decision theorist's thought that preferences are essentially non-graded, whereas desires are essentially graded. Indeed, this assumption is precisely the reason Pollock's objection works. The computational explosion Pollock

¹¹ Pollock, *Thinking about Acting*, *op. cit.*, pp. 22-27.

¹² *Ibid.*, p. 26.

describes arises in the following way. If there were just 300 basic kinds of states of affairs that mattered to us, then all other kinds of states of affairs that mattered to us would correspond to conjunctions of these more basic kinds of states. In computing the value of these conjunctions, however, if our basic evaluative states are non-graded, then there will be no straightforward way of deriving the value of the conjunctions from the value of the conjuncts. Instead, to determine facts about the value of the conjunctions, we will simply have to posit more of these evaluative states. To see this, suppose you prefer “basic” outcomes O1 to O2, and O3 to O4. If these preferences are non-graded, then notice that they alone will not determine whether you also prefer, say, (O1&O4) to (O2&O3), or vice versa. That is, since an ordinal ranking simply tells us that O1 ranks above O2, and that O3 ranks above O4, we’re left without enough information to tell how (O1&O4) and (O2&O3)—as well as other related conjunctions—rank with respect to one another. So, for an agent’s preference ranking to guide their decision making, the ranking will have to explicitly include a further set of preferences regarding all outcomes like (O1&O4) and (O2&O3), for all conjunctions of basic kinds of states of affairs that matter to us. As we’ve seen, Pollock calculates that a ranking of this kind could not plausibly be stored in the brain. In contrast, he notes, if our basic conative states were essentially graded, things would be significantly simpler. A cardinal measure allows us to add values, so if the cardinal measure were more basic, a far greater number of evaluative outputs could be derived from far fewer basic evaluative states. Pollock thus concludes that we should take desire, rather than preference, to be the more basic conative state.

More recently, however, Armin Schulz has argued that Pollock’s argument does not ultimately give us reason to reject the preference-first view.¹³ And this, Schulz thinks, is precisely because the argument assumes that preferences are non-graded. According to Schulz, we simply have no reason to think preferences are non-graded—and, once we drop this assumption, we have

¹³ Schulz, “Preferences vs. Desires,” *op. cit.*, pp. 245-49.

no reason to think the preference-first view would require any implausible amount of computation. After all, if preferences are graded, then there will be no issue regarding adding values, and no significant difference here between preference and desire seems to emerge. We can determine your preference regarding (O1&O4) and (O2&O3), for instance, by summing the degree of your (basic) preference for O1 over O2, say it's 10, and the degree of your (basic) preference for O4 over O3, say it's -5, yielding the degree of your preference for (O1&O4) over (O2&O3): 5. So, as long as we abandon the assumption that preferences are essentially non-graded, we can accept the preference-first view while avoiding Pollock's objection. Schulz concludes:

All that Pollock has shown is that fundamental conative states are likely to be graded—non-graded, purely binary conative states run into computational problems. This is an important insight that shores up the commonly made assumption that conative states are graded. However—and this is key for present purposes—it does nothing to settle the debate about the fundamentality of desires and preferences.¹⁴

As I see it, Schulz is right that Pollock's arguments don't ultimately undermine the preference-first approach as a whole. However, three things are worth noting here. First, while Schulz offers little by way of explicit argument for the thought that preferences come in degrees, it's worth mentioning at least one major consideration in favor of this thought. This is simply that it accords with how we naturally think of preferences in everyday life. Recall: the view that preferences are essentially non-graded was adopted only because an ordinal ranking was the most we could reliably introspect. Presumably, though, those of us without any pre-theoretical commitments would never even think to consider this view. After all, as anyone with preferences seems to know, some

¹⁴ *Ibid.*, p. 249.

preferences are clearly stronger (or weaker) than others, just as some desires are clearly stronger (or weaker) than others. For instance, I strongly prefer my sister's health to a piece of chocolate, which I just slightly prefer to a cookie. I strongly prefer getting one million dollars to getting two dollars, which I just slightly prefer to getting one dollar. Introspection reveals that many of our preferences are graded—it's just that we can only introspect an ordinal measure, rather than a precise cardinal measure, of any given preference. Once we drop the demand of introspectability, then, we have no reason to assume that desires are essentially graded and preferences are not. So, we should drop this demand, and accept—alongside Schulz, and in accordance with our pre-theoretic thoughts about preferences—that preferences and desires both come in degrees.

Second, while it's clear that Pollock's argument hinges crucially on the thought that preferences are essentially non-graded whereas desires are essentially graded, his operating under this assumption seems to me entirely apt. His opponents, after all, are the very decision theorists (etc.) who introduced this apparent difference, and it is this proposed difference that motivates the decision theorist's representation theorems, and their adoption of the preference-first view, in the first place. Seen in this way, then, Pollock's argument seems to successfully show that preferences—understood in the way his opponents seem to intend—cannot be our most basic form of conative state. He thus establishes the important conclusion, not just that our basic form of conative state is likely to be graded, but that the standard decision theorist's version of the preference-first view fails.

Finally, Pollock's argument does not just show that this version of the preference-first view fails. It also shows, more crucially, that the very same thing that motivates the widespread acceptance of the preference-first view also leaves the view with unacceptable psychological implications. That is, as we've seen, the thought that preferences are essentially non-graded whereas desires are essentially graded is precisely what motivates the decision theorist's representation theorems and hence their adoption of the preference-first view. Yet this thought is also, given

Pollock's argument, precisely what leads the preference-first view to a computational explosion. So, Pollock's argument shows that the primary motivation for the preference-first view—the reason it is so widely accepted among decision theorists, economists, and others—is entirely undermined. As a theory of the basic conative structure of human minds, the preference-first view receives no support at all from its most important, storied, and widely accepted source of motivation. Despite the view's dominance, we're then left without any clear reason to accept it at all.

Here then is the upshot of our discussion thus far. The preference-first view enjoys pride of place in decision theory, economics, and elsewhere. However, the same thing that motivates the widespread acceptance of this view—namely, the thought that preferences are essentially non-graded, whereas desires are essentially graded—is also the thing that leaves the view with unacceptable psychological implications. It therefore seems that the preference-first view's standing in the debate over the basic conative structure of human minds is unwarranted. Without this key motivation, as well as resulting support from various representation theorems, we're in fact left with little independent reason at all to accept this view. Our main upshots are then: first, that we should abandon the standard decision theorist's claim that preferences are essentially non-graded whereas desires are essentially graded, and instead claim that both attitudes come in degrees; and second, that in light of this, we should not take the preference-first view to have any established advantage over potential competing views.¹⁵

¹⁵ Admittedly, Schulz also goes on to offer his own clever argument in favor of the preference-first view (*ibid.*, pp. 249–51). In particular, he appeals to certain well-known, systematic choice intransitivities, and argues that the preference-first view is better positioned to explain these intransitivities than the desire-first view. While I think his argument deserves a far more thorough explanation and analysis than this, though—and while I in fact have a great deal to say about how we might explain these intransitivities on the account I'll propose—I simply lack the space both to offer my own account and to fully explain and respond to his argument in this paper (as this would require offering a further account of the relationship between desire, preference, and context). I therefore leave further discussion as a focus for future research.

Rather than continue to discredit the preference-first view, I'd like to turn now to a different tack. Specifically, I would like to introduce an alternative approach, which I will call the desire-first view.

3. The desire-first view

We've now established that there is at least one fundamental difference between preference and desire: preferences are essentially comparative, whereas desires are essentially monistic. Beyond this, it is not immediately obvious how preference and desire fundamentally differ, or whether they do at all. But what about their similarities? Which features do preference and desire seem to share?

The answer is that there seems to be a significant overlap in the features of preference and desire. This overlap seems so significant, in fact, that the difference in intentional objects—preferences having two, and desires having just the one—may be the only obvious, uncontroversial fundamental difference between the attitudes. After all, preferences and desires, as perhaps the two most paradigmatic conative attitudes, seem to play very similar roles in human psychology. To take just a few features, it seems entirely natural to attribute to both desires and preferences each of the following: (i) they tend to entail motivation, or dispositions to act so as to make their objects more likely to obtain; (ii) they tend to cause pleasant feelings when they are seen as or imagined to be satisfied, and unpleasant feelings when they are seen as or imagined to be frustrated; (iii) they tend to direct our attention toward things we associate with their objects; and (iv) they tend to have amplified such effects when we have sensory or imaginative representations of things we associate with their objects.¹⁶ So, for instance, both when I desire coffee, and when I prefer coffee to the

¹⁶ See Neil Sinhababu, "The Humean Theory of Motivation Reformulated and Defended," *The Philosophical Review*, CXVIII, 4 (October 2009): 465-500; and Neil Sinhababu, *Humean Nature* (Oxford: Oxford University Press, 2017), for discussion of these features as they relate to desire; and see James Fanciullo, "On Sense and Preference," *Journal of Moral Philosophy*, XIX, 3 (June 2022): 280-302, for discussion of the overlap here with preference.

water that's in front of me, I'll be disposed to act so as to make it more likely that I'll have coffee; I'll tend to experience pleasant feelings when I see myself as having coffee or imagine that I will, and unpleasant feelings when I see myself as not having coffee or imagine that I won't; I'll be apt to attend to things like the mug on my desk, or my feelings of sluggishness; and these effects will be amplified when I see, smell, or imagine drinking, the coffee. It thus seems that preferences and desires equally drive, and explain, (certain) actions, feelings, and changes in attention.¹⁷ My desiring coffee and my preferring coffee seem equally well-suited to explain these further facts about me. This significant overlap also seems to explain why the terms "preference" and "desire" are often used roughly interchangeably, even in philosophical debates.¹⁸ The attitudes can play many of the same psychological roles, and seem to share many of the same features.

Importantly, I don't offer the above features as anything like a set of necessary and sufficient conditions for desiring or preferring, or a full account of the nature of desire or preference. I'm simply noting a striking overlap in our more intuitive, pre-theoretic understandings of preference and desire. In fact, my aim is for my own view to be ecumenical: it should fit naturally with a wide variety of proposed views of preference and desire (as we'll see).¹⁹ But the point here is simply that preferences and desires seem to share a striking number of features. They seem to share so many, in fact, that it would be surprising if this overlap weren't explained by the fact that one of the attitudes partly constituted the other. It would make little sense, after all, if the very same acts (or the like)

¹⁷ Regardless of whether they explain all actions, feelings, and changes of attention, or just some of them, it seems plausible that they will in any case explain the same ones. So if it is claimed, for instance, that moral belief alone can motivate certain moral actions, this will seem to imply that neither preference nor desire is necessary for such action. This is perfectly consistent with what I'm claiming, which is simply that the attitudes generally seem capable of playing the same roles, regardless of what we ultimately take these roles to be.

¹⁸ Here I have in mind especially the debate over the nature of well-being—see, for example, Chris Heathwood, "Desire-Fulfillment Theory," in Guy Fletcher, ed., *The Routledge Handbook of the Philosophy of Well-Being* (New York: Routledge, 2016), pp. 135-47.

¹⁹ Outside of decision theory and economics, it is exceedingly rare for a theory of desire or preference to commit one to any view of the relationship between desire and preference. Thus my own view, which concerns the relationship between desire and preference, should remain agreeable to almost any theorist with an independent view of either attitude. More on this below.

were explained by two irreducible motivating attitudes, when just the one would do.²⁰ So, it stands to reason that one of these attitudes is more basic than, and explains the similar psychological roles of, the other. According to the preference-first view, we've seen, preference is our most basic form of conative attitude, and desire is to be explained in terms of preference. In contrast, according to the

desire-first view: desire is our most basic form of conative attitude, and preference is to be explained in terms of desire.

On this view, all facts about preferences will be determined, at least in part, by facts about desires. Notice, again, that the desire-first view commits us just to a theory of the relationship between preference and desire, and not to any more precise theory of the nature of these attitudes. As we've seen, there is a strong presumption in favor of the rival, preference-first view in the literature. As we've also seen, however, this presumption is ultimately unwarranted. So what reasons, if any, do we have for accepting the desire-first view?

Before offering several such reasons, I'd like to quickly flag the one fundamental difference between preference and desire that I'll be assuming in what follows. This is that preferences are essentially comparative, whereas desires are essentially monistic. Since this is likely the least controversial proposed difference between preference and desire—indeed, I'm aware of no one who rejects it—my arguments should appeal to a wide audience.

²⁰ An alternative to the preference-first and desire-first views is the view that both preference and desire are basic conative attitudes—or, that neither attitude reduces to the other. The current point can also be seen as an objection to this alternative view: why should we posit two irreducible attitude-types, when it seems one by itself could explain just as much?

4. Preference as desire

To motivate the desire-first view, it's worth thinking about the nature of comparisons more generally. Preferences, as I say, are essentially comparative: to prefer is always to prefer one thing over another. Yet an underappreciated feature of preference in this connection is that comparisons in general seem to share a certain evaluative structure. Thus preferences, as comparisons themselves, will presumably share this structure too. To get a better idea of the structure I have in mind, let's consider a few paradigmatic examples of comparisons.

Suppose we have two things, A and B. As a preliminary point: if all we know is that we want to compare the two things, this will not help us much. After all, we cannot compare two things *as such*—we cannot simply or flatly “compare” them. We must know how, or in what way, we are comparing them. That is, we must know the *dimension along which* the two things are being compared, or what Ruth Chang calls the relevant “covering consideration.”²¹ Do we want to know which is larger, faster, louder, or what? Once we know this, we can happily compare A and B in these different respects, and determine that A is larger, faster, or louder than B. Comparisons, then, are always made relative to some covering consideration, or some dimension along which the objects are compared.

These covering considerations provide *standards of comparison*, which we appeal to when making comparisons. This can be illustrated by our examples just above. Determining that A is faster than B involves comparing A and B in terms of how they measure on a certain standard, namely *speed*. Here we compare A and B with respect to speed—it is the covering consideration—and speed is also the standard to which our comparison appeals. Similarly, determining that A is larger or louder than B involves comparing A and B in terms of how they measure on the standards of *size* or *volume*, respectively. Here we compare A and B with respect to size or volume—they are

²¹ Ruth Chang, “The Possibility of Parity,” *Ethics*, CXII, 4 (July 2002): 659-88.

the covering considerations—and size or volume is also the standard to which our comparison appeals. In each case, our comparisons are made by appeal to how the two things measure up on a specific standard. And these standards themselves seem to be non-comparative.

Preferences are also comparisons. So, they presumably share these general features of comparisons. Specifically, when we prefer A to B, it seems our comparison appeals to some standard, relative to which the comparison proceeds. That is, our preference for A over B appeals to how A and B measure up on some specific standard. What is this standard?

Notice first that, regardless of the answer, these considerations already suggest that preferences must be a mental “output” of some kind. After all, if we reach preferences by appealing to some standard, then this means there must be some mental process (namely, the process involved in appealing to the standard) that results in, and that is therefore prior to, the preferences themselves. Since there is a separate standard to which preferences (as comparisons) appeal, in other words, preferences themselves must be comparisons or measures of something more basic. In the same way, since there is a separate standard to which, say, *larger than* claims appeal, these claims themselves must be comparisons or measures of something more basic—namely, size. We appeal to the individual sizes of A and B (say x and y respectively) in order to determine that A is larger than B (given that x is greater than y). Similarly, it seems, there must be some standard, akin to size, to which we appeal in preferring A to B. Accordingly, the fact that preferences are comparisons suggests that they are derived from a more basic evaluative standard, or some more basic evaluative dimension along which they compare. If all that is right, moreover, it suggests that the most fundamental evaluative data stored in the brain does not come in the form of preferences. This is because, again, there must be some further evaluative data—or some further evaluative standard—to which preferences, as comparisons, appeal. Preferences are ultimately comparisons of this further evaluative data.

Here is another way of putting the point. To prefer A to B, there must be some standard to which you appeal. Appealing to this standard seems to involve retrieving explanatorily prior values, which you assign to A and B individually.²² If this were not the case, after all, then it is unclear how you could come to rank A over B at all: your ranking would appeal to some standard, but this appeal would provide you with no information as to the evaluative standing of A and B. So how would you reach the ranking? By way of analogy: to determine that A is larger than B, there must be some standard to which you appeal. Appealing to this standard seems to involve retrieving explanatorily prior values, which you assign to A and B individually. These values, of course, are measures of size: you assign x to A and y to B. If this were not the case—if there were no prior values to which you appealed—then it is again unclear how your comparison of A and B would proceed. After all, you would be appealing to some standard (namely size), but this appeal would provide you with no information as to how A and B measured up on that standard. There would thus be no relevant information to base your comparison on, and so it is unclear how your comparison could proceed. The same seems to go for preference: coming to prefer A over B seems to require the retrieval of values that are assigned to A and B, and that are explanatorily prior to the comparison.

Now, returning to our question—what is this standard to which preferences appeal?—there are of course many potential answers. In the limited space that remains, rather than construct and tear down alternative answers, I'd like to present and illustrate the plausibility of my own answer. Specifically, my own view is this: *preferences are comparisons of the extents to which alternatives are desired*. That is, the further evaluative data—or the further evaluative standard—to which preferences, as comparisons, appeal, is the extent to which given things are desired. A preference for A over B is

²² Of course, as we've seen, we don't seem capable of introspecting precise cardinal measures here, so I don't mean to suggest that these assigned values must be precise. It might instead be, as Schulz suggests, that the measurement of the values is somewhat rough or arbitrary (Schulz, "Preferences vs. Desires," *op. cit.*, p. 241). (And the same might go for measures of size etc.)

thus a comparison of the extents to which A and B are desired, where A is desired more than B. (And *indifference* is the state where neither alternative is desired to a greater extent than the other.)

There is much more to be said about this proposal. Perhaps most pressingly: what do we mean by “desired?” Do we have any reasons for thinking preferences require desires? Does this proposal have any advantages over the preference-first approach? And many more besides. While I admittedly cannot explore all these questions—or even just the more pressing ones—in their entirety here, I’ll at least offer preliminary answers to these more pressing questions in turn, and conclude that the plausibility of the answers gives us reason to be confident in the view.

4.1 What do we mean by “desired?”

As I’ve mentioned, I hope for this desire-first view to be largely ecumenical. This may seem unlikely, as there are, of course, many views of the nature of desire.²³ And there are several analyses of the nature of preference.²⁴ Moreover, matters here might seem to be significantly complicated by the fact that “desire” and “preference” are each sometimes taken to have multiple senses.²⁵ Fortunately, however, I believe the case demonstrating the proposed view’s wide acceptability is fairly straightforward. This is because there is significant overlap in this existing work on preference and desire.

Take, for instance, the view that to desire p is ultimately to be motivated or disposed to bring p about in certain conditions.²⁶ This view shares an obvious relation to the view on which to

²³ See, for example, Peter Railton, “That Obscure Object, Desire,” *Proceedings and Addresses of the American Philosophical Association*, LXXXVI, 2 (November 2012): pp. 22-46; Schroeder, *Three Faces of Desire*, *op. cit.*; Sinhababu, “The Humean Theory,” *op. cit.*; and Michael Smith, “The Humean Theory of Motivation,” *Mind*, XCVI, 381 (January 1987): pp. 36-61; and, for an overview, see Schroeder, “Desire,” *op. cit.*

²⁴ See, for example, Broome, *Ethics out of Economics*, *op. cit.*; Hausman, “Mistakes about Preferences,” *op. cit.*; Hausman, *Preference, Value, Choice, and Welfare*, *op. cit.*; and Ralph Wedgwood, “Must Rational Intentions Maximize Utility?,” *Philosophical Explorations*, XX, sup2 (October 2017): 73-92.

²⁵ See, for example, Chris Heathwood, “Which Desires Are Relevant to Well-Being?,” *Nous*, LIII, 3 (September 2019): 664-88; and Fanciullo, “On Sense and Preference,” *op. cit.*

²⁶ See, for example, Smith, “The Humean Theory of Motivation,” *op. cit.*

prefer p over q is ultimately to be motivated or disposed to choose p when the alternative is q.²⁷

Notice that these are not just possible views—they are major contenders in debates over the nature of desire and preference, respectively. And notice how easily the analysis offered here—according to which preferences are comparisons of the extents to which alternatives are desired—could explain the truth of these two views. If desiring p is a matter of being motivated to bring p about, and preferring p over q is a matter of being motivated to choose p when the alternative is q, then it will seem entirely natural to explain these two facts in terms of the fact that the preference for p over q is a comparison of the extents to which one is motivated to bring p and q about, where one is more motivated to bring about p. Hence, the truth of these two closely related views—if indeed they are true—can be naturally explained by the version of the desire-first view proposed here. Insofar as we find either of these views plausible, the desire-first view will offer us a straightforward, and seemingly quite plausible, answer to the question of how our view of the one attitude relates to views of the other.²⁸

What goes for proposed views of desire and preference also seems to go for proposed senses of each of “desire” and “preference.” Chris Heathwood, for instance, nicely develops the intuitively plausible distinction between what he calls “desire in the genuine-attraction sense,” or

²⁷ See, for example, Broome, *Ethics out of Economics*, *op. cit.*; and Savage, *The Foundations of Statistics*, *op. cit.*

²⁸ Of course, this is just one view of desire, and just one view of preference. And, admittedly, I lack the space to consider many alternative views here. It at least seems worth noting, though, that many other standard views of desire (see, for example, Schroeder, “Desire,” *op. cit.*) will seem to fit naturally with the proposed desire-first view—and, in fact, these standard views may even yield plausible analyses of preference, when paired with our desire-first view. To take just one example—and to provide just a bit more proof of concept—we can briefly look at Schroeder’s own, well-known analysis of desire (Schroeder, *Three Faces of Desire*, *op. cit.*). On Schroeder’s view, to desire is essentially to use the capacity to represent something as a reward. Hence, together with our desire-first view, Schroeder’s view can be extended to claim that to prefer is to be such that one represents one alternative as a reward to a greater extent than one represents the other alternative as a reward. Preference, then, is a kind of comparison of rewards (or punishments—see especially Schroeder, *Three Faces of Desire*, *op. cit.*, chapter 5). On its face, this view of preference seems to me roughly as plausible as Schroeder’s view of desire—which is to say, it seems to me plausible enough to be one of the standard views of preference. Of course, all of this is very quick, and much more would have to be said to back up any apparent plausibility here. But the point here is again just that, insofar as we find Schroeder’s view of desire plausible, the proposed desire-first view seems to yield a roughly equally plausible view of preference. And, again, I’m confident that something similar will be true of other leading theories.

genuine desire, and “desire in the merely behavioral sense,” or *behavioral* desire.²⁹ To illustrate: sometimes when we desire things, we’re not just motivated to bring the things about, but are genuinely attracted to or enthused by the things. Think, for instance, of desiring to spend time on one’s favorite hobby, or to see one’s partner after a long time apart. In these cases, it’s not just that we’re disposed to make the thing more likely, but that we find the thing attractive, alluring, or appealing: we have what Heathwood calls a *genuine* desire for the thing. In contrast, other times when we desire things, we’re merely motivated to bring the thing about. Think, for instance, of doing tedious and meaningless work, or of going to the dentist despite loathing the thought of doing so. So long as we do these things, it seems we must have in some sense been motivated to do them, and it thus seems we must have in the same sense desired to do them. In these cases, while we’re disposed to act, there is nothing we find attractive or appealing about the things—nothing we view with enthusiasm, excitement, or gusto. And here we have what Heathwood calls a mere *behavioral* desire for the things. This distinction seems deeply plausible and, in fact, has been proposed and endorsed (albeit in slightly different terms) by a variety of philosophers.³⁰ Perhaps unsurprisingly, a very similar distinction has been noticed in the case of preference. At least in one case, in fact, it has been argued that the distinction can be understood in Heathwood’s own terms—or, that “preference” also admits of both genuine-attraction and behavioral senses.³¹ More generally, though, several philosophers have noted a distinction between “preferring” in the sense of *favoring* (or *liking more*), and “preferring” in the sense of *choosing*.³² In one sense, for instance, I prefer staying home to

²⁹ Heathwood, “Which Desires Are Relevant to Well-Being?,” *op. cit.*

³⁰ Examples include David Lewis, “Desire as Belief,” *Mind*, XCVII, 387 (July 1988): 323-32; and Derek Parfit, *On What Matters*, Volume 1 (Oxford: Oxford University Press, 2011). For a much more exhaustive list, see Heathwood, “Which Desires Are Relevant to Well-Being?,” *op. cit.*

³¹ See Fanciullo, “On Sense and Preference,” *op. cit.*

³² See, for example, Hausman, “Mistakes about Preferences,” *op. cit.*; and Wedgwood, “Must Rational Intentions Maximize Utility?,” *op. cit.* In fact, Hausman and Wedgwood both appeal to dictionary definitions for “preference” and “prefer,” and, as Wedgwood puts it, “the dictionaries all agree” that the terms have these two main senses (Hausman, “Mistakes about Preferences,” *op. cit.*, p. 5; Wedgwood, “Must Rational Intentions Maximize Utility?,” *op. cit.*, pp. 78-79).

going to the dentist. As long as I go to the dentist, however, there is another sense in which I must prefer going to staying home. This is explained by the fact that staying home is what I *favor* or *like more*, whereas going to the dentist is what I *choose*. Importantly, notice that there is a clear overlap between this distinction and Heathwood's: staying home is both what I see as more *genuinely attractive* and what I *favor*, whereas going to the dentist is both what I am more *behaviorally disposed* to choose and what I do *choose*. And, again, the desire-first view offered here provides a natural explanation of this overlap. On this view, preferences are comparisons of the extents to which alternatives are desired. So, we can claim: to prefer A to B in the sense of *favoring* A over B is to *genuinely* desire A to a greater extent than B; and to prefer A to B in the sense of *choosing* A over B is to *behaviorally* desire A to a greater extent than B. Accordingly, the truth of these two, independently proposed and independently plausible distinctions can be naturally explained by our desire-first view. Insofar as we find either of these distinctions plausible, the desire-first view will offer us a straightforward, and seemingly quite plausible, answer to the question of how our view of the one distinction relates to views of the other.

Finally, notice that the desire-first view seems to in each case offer a plausible explanation as to why, as we've seen, both desire and preference seem to come in degrees. If preferences are comparisons of the extents to which alternatives are desired, and these desires are themselves graded, then it will be no wonder that the preferences comparing them are graded as well. On this view, the fact that I strongly prefer my sister's health to a piece of chocolate, which I just slightly prefer to a cookie, can be easily explained by the fact that my desire for my sister's health is strong, whereas my desire for a piece of chocolate is weak, and my desire for a cookie is weaker still (again, whatever our understanding of "desire" is here). Facts about strengths of preference can be easily

explained in terms of facts about strengths of desires—and leading views of each attitude seem to fit well with this form of explanation.³³

While I don't take my discussion here to be in any way exhaustive, I do take it to go some way toward illustrating that the desire-first view offered here should prove accommodating to a wide variety of extant views regarding preference and desire.³⁴ Moreover, it seems that it will also shed light on the explanatory relation between these views, and that we have reason to believe, given the explanations we've seen thus far, that the explanations it provides will in each case be straightforward and plausible. So, to answer the question with which we began—what do we mean

³³ The fact that both desire and preference come in degrees raises an important question regarding introspectability, on the current account. As we've seen, it seems we cannot introspect a precise cardinal measure of our desires, yet it also seems our preferences provide precise, introspectable ordinal rankings. Hence if preferences are just comparisons of the extents to which alternatives are desired, the question is, how might we explain this apparent difference in the introspectability of the two attitudes? To put the question slightly differently: if we cannot introspect a precise cardinal measure of our desires, and preferences are just comparisons of these desires, how do preferences suddenly provide introspectable rankings?

Two things can be said in response. The first is simply that, while we cannot introspect a precise cardinal measure of our desires, we often can introspect a rough cardinal measure of our desires. And, when our aim is only to reach a comparative ranking of two desires, this rough measure is typically enough for us to reach such a ranking. I cannot, for example, introspect a precise cardinal measure either of my desire that my sister is healthy or of my desire that I get a piece of chocolate (or indeed, plausibly, of any of my desires). Still, I clearly can introspect a rough cardinal measure of these desires, and this rough measure can allow me to determine that I rank my sister's being healthy over my getting a piece of chocolate. In many cases, then, it seems the rough introspectability of desire will be enough to ensure the precise introspectability of ordinal rankings.

Second, it is worth noting that there are accounts of the introspectability of desire that fit well with this proposed picture. On Lauren Ashwell's account of desire introspection, for example, "Wanting things makes you see them in a certain light, and this is how you introspectively know what you want" (Lauren Ashwell, "Deep, Dark...or Transparent? Knowing our Desires," *Philosophical Studies*, CLXV, 1 (August 2013): pp. 245-56, at p. 255). While, as I've stressed, I aim to remain largely noncommittal as to competing views of the nature of desire, it seems worth noting that Ashwell's account fits well with our earlier, pre-theoretic thoughts about the characteristic features of desire. For example, desires typically cause pleasant feelings when they are seen as or imagined to be satisfied, and typically direct our attention toward things we associate with their objects. These features plausibly explain Ashwell's pleasant "light" (Sinhababu, *Humean Nature*, *op. cit.*, p. 32). More importantly, on Ashwell's account, we can then say that our inability to introspect a precise cardinal measure of our desires is explained by the fact that we cannot introspect a precise cardinal measure of this pleasant "light." Still, since we can introspect that some things are better "lit" than others (or, less metaphorically, that they are more pleasant to think about than others), we can additionally say, this is enough for us to introspect a precise, ordinal ranking of any two desires. Thus, Ashwell's account seems to bear out the explanation of introspectability offered just above. Of course, again, this is just one view of desire introspection (and, admittedly, it may involve substantive assumptions about the nature of desire). But I think it is worth noting, again, that an existing view seems to fit well with the account proposed here. Thanks to an anonymous referee for pressing me on this.

³⁴ A further issue here is that our everyday conception of "desire" in fact seems to capture two distinct psychological phenomena: *positive desire*—a pro-attitude that is fundamentally "for" things—and *aversion*—a corresponding con-attitude that is fundamentally "against" things. See James Fanciullo, "Desire, Aversion, and Welfare," *Analysis*, (forthcoming); Schroeder, *Three Faces of Desire*, *op. cit.*; Sinhababu, "The Humean Theory," *op. cit.*; and Sinhababu, *Humean Nature*, *op. cit.* While I lack the space to fully discuss this distinction and its implications for my account here, I'll at least note that as it figures in the desire-first view, "desire" can be understood as encompassing both positive desire and aversion.

by “desired?”—I leave it largely unsettled. In particular, I remain largely agnostic as to what it ultimately is to desire something, and thus what it ultimately is to prefer something. What I am committed to, though, is that preferences can plausibly be viewed as comparisons of the extents to which alternatives are desired—however “desired” is ultimately to be understood.

4.2 Why think preferences require desires?

If preferences are comparisons of the extents to which alternatives are desired, then preferences are partly determined by desires, and preferences thus require desires. But what reasons, if any, do we have for thinking that preferring is partly a matter of desiring?

Intuition suggests that we can desire without preferring, but we cannot prefer without desiring.³⁵ Regarding the latter point: notice that if you have no desire for anything related to either A or B, it seems plausible that you cannot have a preference between them. If, for instance, you have no desire for anything relating to the fact that $2+2=4$ —if there is absolutely nothing relating to this fact that you want or that moves you—and you similarly have no desire for anything relating to the fact that $3+3=6$, then it seems you cannot have a preference concerning these two facts. The lack of any desires, not just for the two alternatives themselves, but for absolutely anything relating to the two alternatives, seems to rule out having a preference between the alternatives. Of course, you can have a preference between A and B when you have no desire for anything relating to one of A and B. But this can only be the case, it seems, when you have a desire for something relating to the other of A and B. For instance, I might prefer my sister’s health to my learning some inane piece of information, even if I have no desire for anything relating to hearing this inane information, since I have a desire regarding my sister’s health. As soon as we assume I have no desire for either

³⁵ The points in this paragraph are inspired by (and follow) arguments presented by Fanciullo, “On Sense and Preference,” *op. cit.*, pp. 287-88.

alternative, though, it suddenly seems plausible to think I cannot have a preference between them. These simple cases then suggest that we cannot prefer without having some relevant desire.³⁶ By contrast, regarding the first intuition mentioned above, it seems plausible that we can desire A or B without having any preference between A and B. If, for instance, a subject at a coffee shop desires to drink the coffee in front of them, and also desires to play with their children later, it seems entirely natural to think that they may nevertheless have no preference between these two things. After all, they may never have been prompted to compare these two things, or to see them as relevant alternatives. As soon as they're informed that drinking the coffee will (somehow!) prevent them from playing with their children, however, they will be forced to compare their newly conflicting desires, and it will suddenly seem plausible to attribute a preference to them. It therefore seems plausible that we can desire without preferring, but we cannot prefer without desiring.

Note also that, given it seems plausible that one of these attitudes will ultimately be explained in terms of the other, considerations of simplicity seem to suggest that desire will be the more basic attitude. This is simply because desires are monistic, whereas preferences are comparative. If one of the attitudes is to be explained in terms of the other, after all, then it seems likelier that the attitude with the single intentional object would explain the attitude with the two

³⁶ One might object here by appealing to cases where one must form a preference between two bad options. Suppose, for example, that you must choose either to be tortured for ten seconds, or else to be tortured for ten hours. Surely you prefer being tortured for ten seconds. Still, it may seem inapt to say you *desire* to be tortured for ten seconds—it's not something you actually want. Is this then a case of a preference without a relevant desire? It is not. This is because, as mentioned above (fn. 34), "desire" should be understood to encompass both positive desire and aversion. While you do not positively desire either to be tortured for ten seconds or to be tortured for ten hours, I assume, you are averse to both these things. And, indeed, you are more averse to being tortured for ten hours. This is not, then, a case of a preference without a relevant desire, since you in fact have two relevant desires. It's just that these desires are aversions. It is worth noting, in addition, how this analysis fits with the view proposed here. On this view, a preference for A over B is a comparison of the extents to which A and B are desired, where A is desired more than B. In cases where you are averse to both alternatives, we can say the less aversive option is "desired more." Thus when you are averse to both A and B, and your aversion to B is stronger, we can say you desire A more than B. This fits with how we naturally talk: if you're averse both to being tortured for ten seconds and to being tortured for ten hours, and these are your only two options, it seems natural to say that what you desire more is being tortured for ten seconds. Moreover, this understanding of "desiring more" in the case of aversion allows us to stick with our original formulation of the proposed view: a preference for A over B is a comparison of the extents to which A and B are desired, where A is desired more than B. Thanks to an anonymous referee for pressing me on this.

intentional objects, rather than vice versa. Of course, it's possible that the explanation goes the other way—this simple point is far from decisive. But given we have a view that is independently plausible, and that also provides the simpler kind of explanation we'd expect, it seems we have all the more reason to think that preference is explained in terms of desire, and indeed that it is explained in the way the proposed view suggests.

Again, I do not take these brief considerations to be decisive. But they do at least give some support to the idea that preference requires desire. As it happens, they also help to answer our final question, which I'll turn to now.

4.3 Why prefer the desire-first view?

While I lack the space to consider alternative proposals in full here, it's worth highlighting both the simplicity and the explanatory power of the account I'm offering by briefly comparing it with the preference-first approach. We've just seen a variety of things that the desire-first view seems to explain with ease. The preference-first view, of course, will have to explain the very same things. This includes explaining how it could be that preferences seem to be comparisons of some more basic evaluative data, if preferences are indeed our most basic conative states. It also includes explaining the striking overlap in the features of desire and preference. As I see it, however, these explanations are unlikely to be nearly as simple or plausible as those provided by our desire-first view—and this illustrates just how natural the latter view is.

Take, as just one of the things the preference-first view will have to explain, the connection we've seen between (what, for simplicity, we can call) the “genuine-attraction” senses of both “desire” and “preference.” On the current account, this surprising overlap is explained in terms of the fact that genuine preferences are comparisons of the extents to which alternatives are genuinely desired, where one alternative is genuinely desired more than the other. This explanation seems as

plausible as it is simple. In contrast, we might wonder, how could a version of the preference-first view explain this striking similarity in preference and desire? Presumably, proponents of the view will claim that one's genuine desire to, say, play basketball is ultimately to be explained in terms of one's genuine preferences between different alternatives. But which alternatives are these, and how does this explanation go? Frankly, I can't think of any straightforward answer. Is one's genuine desire to play basketball explained in part in terms of one's genuine preference for playing basketball over playing football, or (instead?) in terms of one's genuine preference for playing basketball over working, or perhaps (instead?) in terms of one's total set of genuine preferences? I'm unsure. Regardless of the answer, though, it at least seems clear that the explanation offered here will be considerably more complex than the one offered by the desire-first view, which of course need not explain how we derive (genuine) desires from anything else at all.

As this example illustrates, even if a version of the preference-first view is able to explain the same connections between preference and desire that the desire-first view offered here can, the explanations offered by the desire-first view seem likely to be simpler. Of course, I haven't considered a complete version of the preference-first view here, and it might turn out that some such complete version can explain just as much as the proposed desire-first view, and do so in as simple of terms. Until that view is offered, though, the advantages we've seen here at least seem to establish a presumption in favor of the current desire-first view.

5. Upshots and conclusions

The preference-first view is essentially treated as orthodoxy among standard decision theorists, economists, many philosophers, and others. As we've seen, however, the primary motivation for adopting this view—motivation that explains the view's wide acceptance and huge influence in these fields—also leaves the view with unacceptable psychological implications. As a theory of the basic

conative structure of human minds, the standard preference-first view cannot be correct. This alone is significant.

Moreover, we seem to have strong reasons for thinking the preference-first view gets things the wrong way around. As we've seen, there is an alternative view, the desire-first view, which despite being largely ignored, on its face seems no less plausible than the preference-first view. In addition, whereas the standard preference-first view fails, we've seen that there is a plausible version of the desire-first view available. This view is motivated by the thought that comparisons in general proceed by appeal to a non-comparative standard or dimension along which alternatives are compared. Thus preferences, as comparisons, presumably appeal to some such standard too. We can then offer a natural explanation here, on the desire-first view, by claiming that the standard to which preferences appeal is the standard of desire: preferences are comparisons of the extents to which alternatives are desired, where one alternative is desired to a greater extent than the other. We've seen that this view fits naturally with several important views of preference and desire, and even offers a unifying explanation of a striking overlap in extant discussions of preference and desire.

The desire-first view offered here thus has a good deal of intuitive appeal, both on its face and in its implications. In contrast, we're left without any similarly plausible version of the preference-first view. I conclude, therefore, that the standard view of human evaluative psychology should be revised.³⁷ Compared to the preference-first view, the desire-first view seems to give us

³⁷ Of course, if I'm right about this, there remain many questions to be answered, and many lines of future research to pursue. To name just a few, we might consider: whether the truth of the desire-first view casts doubt on the usefulness of the decision theorist's representation theorems; whether the desire-first view can meet the challenge (mentioned above) of explaining certain systematic choice intransitivities (Schulz, "Preferences vs. Desires," *op. cit.*); or whether the desire-first view might shed light on the economist's problem of explaining why preferences seem to depend in significant and systematic ways on seemingly irrelevant contextual factors (see, for example, Sarah Lichtenstein and Paul Slovic, eds., *The Construction of Preference* (New York: Cambridge University Press, 2006)). While I think there is much more to be said about each of these questions—especially regarding the relations between desire, preference, intransitive choice, and context—the aim of this paper was just to introduce and establish the desire-first view as an important, live view. Having made my best case for this, I leave discussion of these other important questions for another time.

more of what we want in our theory of human evaluative psychology—and so, given that the two theories conflict, it seems the desire-first view is to be preferred.³⁸

³⁸ Many thanks to two anonymous referees for helpful comments. This work was supported by the Research Grants Council of Hong Kong [grant no. 23603223].

References

- Ashwell, Lauren. 2013. "Deep, Dark...or Transparent? Knowing Our Desires." *Philosophical Studies* 165 (1): 245-256. doi: 10.1007/s11098-012-9950-3
- Bermúdez, José Luis. 2009. *Decision Theory and Rationality*. Oxford: Oxford University Press.
- Broome, John. 1999. *Ethics out of Economics*. Cambridge: Cambridge University Press.
- Chang, Ruth. 2002. "The Possibility of Parity." *Ethics* 112 (4): 659-688. doi: 10.1086/339673
- Fanciullo, James. 2022. "On Sense and Preference." *Journal of Moral Philosophy* 19 (3): 280-302. doi: 10.1163/17455243-20213579
- Forthcoming. "Desire, Aversion, and Welfare." *Analysis*. doi: 10.1093/analys/anae101
- Hausman, Daniel M. 2011. "Mistakes about Preferences in the Social Sciences." *Philosophy of the Social Sciences* 41 (1): 3-25. doi: 10.1177/0048393110387885
- 2012. *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.
- Heathwood, Chris. 2016. "Desire-Fulfillment Theory." In *The Routledge Handbook of the Philosophy of Well-Being*, edited by Guy Fletcher. Routledge.
- 2019. "Which Desires Are Relevant to Well-Being?" *Noûs* 53 (3): 664-688. doi: 10.1111/nous.12232
- Jackson, Elizabeth G. 2020. "The Relationship Between Belief and Credence." *Philosophy Compass* 15 (6): e12668. doi: 10.1111/phc3.12668
- Lewis, David. 1988. "Desire as Belief." *Mind* 97 (387): 323-332.
- Lichtenstein, Sarah, and Paul Slovic, eds. 2006. *The Construction of Preference*. New York: Cambridge University Press.
- Parfit, Derek. 2011. *On What Matters. Vol. 1*. Oxford: Oxford University Press.
- Pollock, John L. 2006. *Thinking About Acting: Logical Foundations for Rational Decision Making*. New York: Oxford University Press.

- Railton, Peter. 2012. "That Obscure Object, Desire." *Proceedings and Addresses of the American Philosophical Association* 86 (2): 22-46.
- Ramsey, Frank P. 1926. "Truth and Probability." In *The Foundations of Mathematics and other Logical Essays*, edited by R. B. Braithwaite. London: Kegan, Paul, Trench, Trubner & Co.
- Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley.
- Schroeder, Timothy. 2004. *Three Faces of Desire*. New York: Oxford University Press.
- 2010. "Desire and Pleasure in John Pollock's *Thinking about Acting*." *Philosophical Studies* 148 (3): 447-454. doi: 10.1007/s11098-009-9491-6
- 2020. "Desire." In *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/sum2020/entries/desire/>
- Schulz, Armin W. 2015. "Preferences vs. Desires: Debating the Fundamental Structure of Conative States." *Economics and Philosophy* 31 (2): 239-257. doi: 10.1017/S0266267115000115
- Sinhababu, Neil. 2009. "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118 (4): 465-500. doi: 10.1215/00318108-2009-015
- 2017. *Humean Nature*. Oxford: Oxford University Press.
- Smith, Michael. 1987. "The Humean Theory of Motivation." *Mind* 96 (381): 36-61.
- Steele, Katie, and H. Orri Stefánsson. 2020. "Decision Theory." In *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2020/entries/decision-theory/>
- von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. New York: Wiley.
- Wedgwood, Ralph. 2017. "Must Rational Intentions Maximize Utility?" *Philosophical Explorations* 20 (sup2): 73-92. doi: 10.1080/13869795.2017.1356352