

Wanting at Will

Ben Holguín

February 2026

"In things which are absolutely indifferent there can be no choice and consequently no option or will, since choice must have some reason or principle."

–Leibniz, *Theodicy*

"Sensibility does not necessitate its action. There is in man a power of self-determination, independently of any coercion through sensuous impulses."

–Kant, *Critique of Pure Reason*

1 Introduction

Schopenhauer once said that man is free to do as he wills, but not to want as he wills.¹ Here I take Schopenhauer to be rejecting *bouletic voluntarism*, the thesis that human beings can form their desires at will.

So understood, Schopenhauer's dictum is a bit of contemporary philosophical orthodoxy; reflection on paradigm cases of desire makes it easy to see why. Perhaps it's up to me whether I satisfy my desire to snooze an extra ten minutes. But there doesn't seem to be any interesting sense in which it's up to me whether I have that desire in the first place. You can sensibly ask me why I chose to *indulge* my desire to snooze; but you can't sensibly ask me why I chose to *desire* to snooze. After all, if it were a matter of my choosing, why would I allow myself to be plagued by such yearning?

This kind of case is not exceptional. Introspect and I imagine you'll find few if any desires that you would naturally describe as having been arrived at voluntarily. Maybe you'll find desires that are the *byproducts* of other things you did voluntarily: for example, a desire for nicotine brought on by a voluntary decision to take up smoking. But such cases do not undermine the intended interpretation of Schopenhauer's dictum: voluntarily doing something that causes a change in one's desires is one thing; voluntarily changing one's desires is another.

Nonetheless, I believe that Schopenhauer's dictum is wrong. Ordinary human beings can and often do want at will. This is not because, contrary to appearances, desires like my desire to snooze an extra ten minutes are typically formed voluntarily. (I don't believe that.) Nor is it because we can voluntarily cause ourselves to have different desires, say by taking up smoking.

¹ Or at least sayings like this are commonly attributed to him. The original German is 'Der Mensch kann tun was er will; er kann aber nicht wollen was er will' (Schopenhauer 1960).

(I *do* believe that; it's just beside the point.) The reason I believe Schopenhauer is wrong is because I believe: (i) we can and often do voluntarily decide what we want when weighing between options that are equally desirable, as in the case of Buridan's ass; and (ii) we can and often do voluntarily decide how discriminating our desires are when weighing between options that are almost (but not quite) equally desirable.

Most of this paper (§§2-11) is devoted to explaining and defending these ideas. But I will also take the time to explore a number of their consequences (§§12-14), especially in connection to questions concerning the relationship between desire and action.

2 'Wanting'

I'll start with some preliminary points about the subject matter and methodology of this paper.

In asking whether human beings can want at will, my interest is in the *ordinary* notion of wanting—that is: whatever attitude is picked out by 'want' in natural language sentences of the form '*S* wants *p*'. I will not be making any prior assumptions about this attitude's theoretical roles. Instead, I will be taking claims involving 'want' (and its cognates) more or less at face value—assuming that their felicity implies their truth, that the felicity of their denials implies their falsehood, and so on—with the aim of getting clear on the picture of the mind given to us by such judgments.

On occasion I'll consider certain error-theories, but for the most part I will simply set aside any thoroughgoing skepticism about the probative force of linguistic intuition with respect to the nature of ordinary desire. To be clear, though: despite the fact that much of the argumentation of this paper will center on claims about what agents can or cannot be felicitously *said* to want, the subject matter of the paper is not the word 'want', but the nature and norms of wanting itself. The linguistic judgments are merely meant to serve as windows into the relevant states of mind.

I will use 'want' and 'desire' interchangeably (as I have so far), treating '*S* wants *p*' as synonymous with '*S* desires *p*', '*p* is one of *S*'s desires', and so on. It is questionable whether the ordinary notions of 'want' and 'desire' are genuine synonyms (Davis 1984). But 'desire' is readily nominalized in a way that 'want' is not, making it convenient to treat the two expressions as if they were semantically equivalent. To the extent that this is a distortion, the official target of this paper is wanting, not desire. But from here on out I will ignore this possibility and assume that the question of whether agents can want at will is one and the same as the question of whether agents can desire at will.

I will assume that desire (wanting) is a propositional attitude, and so will be interpreting schematic desire-reports like '*S* wants *p*' as expressing a relation between a subject and a proposition. This is really just for convenience. Ordinary desire reports typically take infinitival ('I want to snooze') and nominal complements ('I want a nap') rather than that-clauses ('I want that I snooze'). But it's easy enough to translate between the two ('I want to snooze' ↔ 'I want that I snooze'). And to the extent that these translations fail to preserve synonymy (Montague 2007), the differences shouldn't be germane to any of the substantive claims I defend in this paper.

In addition to desire, I will also be discussing *preference*. Here too I am interested in the ordinary notion—i.e., the attitude expressed by ‘prefers’ in sentences of the form ‘S prefers *p* to *q*’. I won’t be defending a theory of (ordinary) preference in this paper. Nor will I try to stake out the exact relationship between desire and (ordinary) preference. But I will assume that the two attitudes are connected in certain straightforward ways. For example, I will assume that, generally speaking, you prefer things you want to things you fail to want (i.e., if you want *p* and don’t want *q*, then you prefer *p* to *q*); and that, generally speaking, if you fail to have a preference between two things, then either they are both things you want or both things you fail to want (i.e., if you are indifferent between *p* and *q*, then either both *p* and *q* are compatible with what you want or neither is). In virtue of these sorts of penumbral connections, my arguments for voluntarism about desire should transpose cleanly into arguments for voluntarism about preference.²

Finally, I will be taking for granted that ‘want’ and its cognate notations are highly context-sensitive. Arguments and examples in support of this idea are legion.³ Here’s one for the sake of concreteness (we’ll see many others over the course of the paper): a person can at one point say something like ‘Of course I don’t want to pay my taxes; I’d rather spend the money on frivolities’, and then at another point say something like ‘Of course I want to pay my taxes; I don’t want to go to jail’. Supposing the person says both things, does it follow that they have incoherent desires, or that they underwent some kind of change of mind between the two speeches? I don’t see why it would. For a much simpler explanation of how these apparently contradictory desire reports could be true at once is that ‘want’ is context-sensitive—i.e., that it can express different relations on different occasions of use—with the natural (true) interpretations of the two speeches involving different resolutions of this context-sensitivity. Of course, anyone who favors this sort of view faces residual questions about the parameters of context-sensitivity (ordering sources? alternative sets? expected utility thresholds?), as well as meta-semantic questions about how and why those parameters get set in the way they do on different occasions of use. But we can be confident ‘want’ is context-sensitive even if we’re in the dark about the answers to these difficult questions.

All that said: despite the fact that I take ‘want’ to be highly-context sensitive, I will often be ignoring issues related to its context-sensitivity, and will tend to use rather than mention ‘want’ over the course of the discussion. This is in service of avoiding what would otherwise be an oppressive amount of meta-linguistic ascent. But I also believe this simplifying choice

² There is a substantial literature on the question of the relationship between desire and “preference”—see, e.g., [Luce & Raiffa \(1957\)](#), [Dreier \(1996\)](#), [Pollock \(2006\)](#), [Bermúdez \(2009\)](#), [Schroeder \(2010\)](#), [Schulz \(2015\)](#), [Barrett \(2019\)](#), [Blumberg \(2022a\)](#), and [Fanciullo \(2022, forthcomingb\)](#) for discussion. However, I put ‘preference’ in square quotes because with the exception of [Blumberg \(2022a\)](#) (see especially his §3) and [Fanciullo \(forthcomingb\)](#), most of the contributions to this literature seem not to be explicitly concerned with the relationship between (ordinary) desire and *ordinary* preference (i.e., the attitude picked out by the natural language expression ‘prefers’). Instead, they seem to be concerned with the relationship between desire and a quasi-theoretical notion of preference, which is typically operationalized in terms of an agent’s dispositions to act or choose (and then employed in the construction of a decision-theoretic apparatus). My discussion, by contrast, will focus on the ordinary notion of preference, and will for the most part sidestep questions about its connections to these more theoretical notions. For more on some of the interpretive and methodological issues involved here, see, e.g., [Lewis \(1974\)](#), [Pettit \(1991\)](#), [Hausman \(2011\)](#), [Dietrich & List \(2016\)](#), [Guala \(2019\)](#), [Thoma \(2020\)](#), [Beck \(forthcoming\)](#), and [Fanciullo \(forthcominga\)](#).

³ For helpful discussion of the varieties of context-sensitivity in desire ascriptions—as well as the theoretical consequences of this context-sensitivity—see, e.g., [Davis \(1984\)](#), [Heim \(1992\)](#), [von Fintel \(1999\)](#), [Rubinstein \(2017\)](#), [Blumberg & Holguín \(2019\)](#), [Jerzak \(2019\)](#), [Blumberg & Hawthorne \(2021, 2022\)](#), [Phillips-Brown \(2021, 2025, forthcoming\)](#), [Blumberg \(2022a\)](#).

is substantively harmless, since I believe the kinds of judgments that tell in favor of bouletic voluntarism can be replicated across variations in the relevant parameters of context-sensitivity (whatever they may be). In fact, I believe bouletic voluntarism holds on *all* ways of resolving the context-sensitivity of ‘want’, and thus that the truth of ‘Human beings can and often do want at will’ is not sensitive to context.⁴ But if this universal generalization turns out to be wrong, I am happy to settle for the more conservative claim that bouletic voluntarism holds of the attitudes expressed by ‘want’ on some of its most natural interpretations.

3 ‘At will’

So much for what I mean by ‘want’. What about ‘at will’?

3.1 On doing at will

As a vague first pass, doing something at will involves doing it intentionally and directly, without mediation by means of one’s environment or some other action.⁵ The relevant concept could just as easily be expressed by any of: ‘making up one’s mind’, ‘deciding’, ‘choosing’, ‘picking’, ‘selecting’, or ‘exercising direct agential control’.⁶ But officially I’ll be understanding the notion by ostension.

Mental activities provide a number of paradigm cases of acts of will. When you imagine a scene, think through an argument, or attend to select features of your present conscious experience, it is often by an act of the will. Not always, of course; any of these things can “just happen”—as in the case of intrusive thoughts, sudden environmental stimuli, and so on (Strawson 2003). And unsurprisingly there can be cases where it seems indeterminate whether the relevant mental activity ought to be described as having been done at will versus having occurred by no choice of one’s own.⁷ But that’s all perfectly consistent with the fact that what we imagine, think about, or attend to can be and often is under our voluntary control.

Further paradigm cases of acts of will can be found at the boundary of the mental and the corporeal. Take subvocalized speech. Both whether you’re subvocalizing and, if so, what you happen to be saying to yourself, are often determined at will (Chomsky 1966). Likewise with the facts about what we *try* to do.⁸ You can try at will to lift a finger, picture an ouroboros, or move the moon with your mind. Indeed, in some cases of trying—say when you’re unknowingly

⁴ To be clear, the claim that bouletic voluntarism holds on all ways of resolving the context-sensitivity of ‘want’ is not equivalent to the claim that every *kind* of desire is such that we can and often do form desires *of that kind* at will. *Desires not formed at will* constitute a trivial counterexample. Perhaps “appetitive” desires, in the sense of Davis (1984), or “unmotivated desires”, in the sense of Nagel (1970, p. 29), constitute more interesting classes of counterexamples as well. But these examples are no threat to the context-invariant truth of bouletic voluntarism, since (as far as I can tell) there are no interpretations of ‘want’ on which (e.g.) ‘S wants p’ is synonymous with ‘S appetitively wants p’ and/or ‘S unmotivatedly wants p’.

⁵ For more on the relevant notion of directness, see, e.g., Danto (1965), Alston (1988), Shemmer (2005, §3), and Mele (2017).

⁶ Some philosophers have argued for the importance of drawing a distinction between *picking* and *choosing*, where (roughly) “picking” is a kind of arbitrary selection between options, while “choosing” is non-arbitrary selection (Ullmann-Margalit & Morgenbesser 1977). So understood, both picking and choosing are acts of the will. As it turns out, the cases that motivate the distinction will underly one of the two main arguments I’ll be giving for bouletic voluntarism (§6.)

⁷ See Levy (forthcoming) for relevant discussion of this exact issue in the case of mind wandering.

⁸ See, e.g., O’Shaughnessy (1973, 1980), Pietroski (2000), Holguín & Lederman (forthcoming).

paralyzed or attempting telekinesis—it may be that the relevant trying is nothing over and above the performance of an act of the will.

The examples of speech and trying are also useful in bringing out the distinction between doing something at will and doing something *consciously*. As you subvocalize, there can be many points where you fail to consciously attend to your particular choice of words; but that doesn't stop them from being just that: a choice of words. Likewise with trying: each of the letters in this sentence is such that I tried to type it, and each such typing event was partly the result of an act of the will, but few of the letters were such that I was consciously aware that I was trying to type them. Doing something at will does not entail being consciously aware of what one is doing.

So far I've only mentioned *activities* we perform at will. But there are also *states* we enter (or withdraw from) at will. Consider intention. Over the course of a morning you can go from intending to go to the gym that afternoon, to having no intention either way, to actively intending not to go, with each transition occurring at will. There are also cases in which it is plausible the will is *essential* to our capacity to form intentions—in particular, cases where we are deliberating between two or more equally desirable courses of action, in the manner of Buridan's ass (more on which shortly).⁹

Our power to choose our intentions is not unlimited, of course. A plausible lesson of Kavka's (1983) toxin puzzle is that there are things we just can't help but fail to intend to do, no matter how much it would be in our interest to have those intentions. But this is consistent with the claim that, often enough, the facts about what we do or do not intend to do are determined by voluntary choices (Shepherd 2018).

The last positive paradigm case I'll mention is *guessing*. Since I'll be discussing voluntarism about guessing in some detail in §4, here I'll be brief and point to one quick example. Suppose I show you a fair coin and ask you to guess how it will land. You might guess heads, you might guess tails.¹⁰ What determines what your guess is? As far as I can tell, it's an act of the will: you simply make up your mind one way or the other.

With these examples in mind, the claim that we can form desires at will is to be understood as the claim that we can form desires like *that*—like the way we sometimes guess, intend, try, subvocalize, and picture. Beyond this, I won't be making any further assumptions about what's involved with willing something: the psychological mechanisms underlying it, its epistemic properties, or what have you. I will, however, close this section *via negativa*, and mention some examples of changes in desire that I will *not* be counting as evidence in favor of bouletic voluntarism.

3.2 On *wanting* at will, in particular

Let's return to one of the examples from the introduction. Suppose I choose to take up smoking, knowing full well that doing so will result in my acquiring a (yet unpossessed) desire for nicotine. In fact, suppose I take up smoking for this very reason: so that I can experience the desire for nicotine, *a la* Frankfurt's (1971)'s addiction therapist. I eventually start smoking and,

⁹ See, e.g., Bratman (1987, 1999), Holton (2009, ch. 1).

¹⁰ Or perhaps you'll simply not have any particular guess. More on this possibility in §4.

predictably, acquire the desire for sweet nicotine. Does it follow that this desire was formed at will? No. All that follows is that I've willed myself to do something that (predictably) caused me to have the desire. Here the connection between my willing and my desire for nicotine is too indirect to count as a case of wanting at will (at least in the sense of interest).¹¹ Analogous points go through for some of the other ways we can exert control over our desires: brain surgery, hypnosis, meditation, techniques of positive visualization, and so on. Though it would not be completely unnatural to describe these sorts of desires as having been formed "at will", the relationship between the willing and the desire is too ballistic to be an instance of the phenomenon in which I'm interested.

Cases of ballistic desire formation aside, there are other putative examples of desiring at will that have been discussed in the literature, but that are different in spirit than the kinds I'll be exploring in this paper.¹² Here I'll just mention one.

Ruth Chang (2009, 2013a,b, 2017, 2020) has argued that when faced with so-called *hard choices*—roughly, decisions between options that are naturally described as incomparable, incommensurable, and/or "on a par", like the choice between a career as a lawyer versus a career as a harpist—we can generate *reasons* for favoring certain options over others at will (cf. Raz 1998, Elson 2021, §4).¹³ For example: when weighing between the two careers, one might simply decide that the "dulcimer tones of the harp" are a reason to pursue a career as a harpist, thereby *making* them a reason to pursue it (Chang 2009, p. 247). In virtue of having these freshly willed reasons, one has new, non-arbitrary grounds for intentionally resolving the hard choices one faces, and to identify with one's subsequent course of action.

Let's suppose Chang is right about this restricted form of voluntarism about reasons. From here it's a short step to an argument for voluntarism about desire.¹⁴ For supposing we can (and presumably often do) desire to do what we have most reason to do, it would seem to follow that we also have the power to generate new preferences at will—viz., the preferences that issue from our recognizing that we (now) have most reason to resolve the hard choice in a certain way. To continue with the example: the person faced with the choice between a career in law and a career as a harpist initially has no preference between the two, but by an act of the will comes to see more value in the harp than in the law, and thereby comes to acquire a preference for harpistry. And perhaps this would be deserving of the label 'wanting at will'.

It would be fine for my purposes if this sort of argument for bouletic voluntarism counted. But I think there are reasons for skepticism. One is that the connection between the relevant act of will and the resulting desire is more ballistic than I'd like it to be. It seems to me more natural to describe such cases as ones in which one's desires simply conform to the freshly willed facts about what one has most reason to do than as ones in which the desires themselves have been willed into existence. But obviously these matters are subtle. The more pressing issue is

¹¹ See, e.g., Williams (1973, ch. 9) and Shemmer (2005, §3) for further discussion of the relationship between willing and "directness".

¹² See, e.g., Harman (1976, 1998), Shemmer (2004, 2005).

¹³ There's a bit of a terminological morass around the phenomenon of hard choices, which for the purposes of this paper we do better to avoid. What's meant to be distinctive about these choices is that they involve comparison between options that seem to be neither better, worse, *nor* equally good as one another, in a way that seems to go beyond its merely being vague which of these relations holds between them. See Andersson & Hsieh (2025) for further discussion.

¹⁴ Note: this is not an argument Chang has made or would endorse, to my knowledge. But see Elson (2021).

that this path to bouletic voluntarism involves a number of controversial theoretical commitments—commitments I’d rather not take on if I don’t have to.¹⁵ So the arguments I will give for bouletic voluntarism will *not* be premised on judgments about hard choices.

This section has attempted to get clearer on what is meant by ‘at will’ in the claim that we can and often do want at will. Instead of giving an explicit analysis of the notion, I’ve tried to anchor it to judgments about certain sorts of mental states and activities, as well as to distinguish it from certain other methods of quasi-voluntary desire formation. The next section will argue in more detail for a claim briefly made above: that we can and often do form *guesses* at will. The discussion of guessing will serve two purposes. First, it will more firmly establish the kind of willing I think is involved in typical cases of wanting at will, beyond the gesturing that’s happened so far. And second, it will serve as a sort of prelude to the paper’s central argument for bouletic voluntarism.

4 Prelude: guessing at will

Suppose again I show you a fair coin and ask you to guess how it will land once flipped. Suppose you’re certain the coin is fair and that it will land either heads or tails (rather than explode in mid-air, land on its side, etc.). And suppose you have no better use of your time than to engage with me. What, then, is your guess?

I see three reasonable answers you might give:

- Heads.
- Tails.
- Either heads or tails / Not on its side / I don’t know.

That is: you might guess heads, you might guess tails, or you might be agnostic, guessing only that the coin will land some way or other.¹⁶

Let’s set aside agnosticism for now; we’ll return to it in a moment. Suppose you take an opinionated stance, and suppose in fact your guess is heads. In virtue of what is this your guess?

As far as I can tell, the answer is very likely to be something along the lines of: it’s what you *decided* to guess.¹⁷ What else could really be going on here? Again, by hypothesis you find both possible outcomes equally likely, so it’s not like your estimation of the probabilities gives you reason to favor one outcome over the other. And, if you’re like me, it’s not like you introspected and noticed that there was a guess there and that it was heads. Rather, you considered the question and then made up your mind that way.

¹⁵ Chief among these is the commitment to failures of comparability between options. For a recent argument against the existence of such failures, see [Dorr et al. \(2023\)](#).

¹⁶ Two clarificatory points. First, as I intend to use the word ‘guess’, we would be asking the same thing if instead the question were ‘How do you think the coin will land?’—cf. [Holguín \(2022\)](#). Second, much of what I say in this section will talk about the kinds of guesses you can reasonably *report* yourself as having. The decision to put things in terms of guess-reports is for vivacity, and is not essential to the argument. The same points would go through *in foro interno*. For further evidence that guesses correspond to genuine mental states, see, e.g., [Horowitz \(2019\)](#), [Dorst & Mandelkern \(2023\)](#), [Quillien et al. \(2023\)](#).

¹⁷ This is to say nothing about the more fundamental psychological mechanisms underlying one’s decision to guess heads (though for some potentially relevant discussion, see, e.g., [Icard 2016, 2021](#)), let alone the neurological mechanisms underlying those psychological mechanisms ([Davidson 1970](#)). The question of interest here is folk psychological, and at the level of folk psychology ‘decision’ seems to me the most natural description of the grounds of the guess (cf. [Holton 2009](#), Ch. 3).

More generally, being asked to guess the result of a fair coin toss is like being in the midst of a doxastic Buridan case: you're faced with two options that are equally (epistemically) good, and you can either remain still (be agnostic) or make up your mind between them (guess heads, guess tails). We can and often do make up our minds this way, breaking the tie whichever way we please. In doing so, we exercise direct voluntary control over a certain aspect of our overall doxastic state.

Tie-breaking is not the only role for the will in guessing. This is because there is nothing that forces you to guess one of heads or tails in particular; you could instead be agnostic, remaining still between the two options, as it were. Suppose this is your stance: you neither guess that the coin will land heads nor guess that the coin will land tails, but instead guess merely that the coin will land some way or other.¹⁸ Once again the question arises: in virtue of what is this your guess? And once again it seems to me that, at least in many ordinary cases, the reason your guess is 'Either heads or tails'—rather than 'Heads' in particular or 'Tails' in particular—is because that's how you've made up your mind.¹⁹ Nothing about the situation makes it inevitable that you have an opinion or that you lack one. It's up to you to decide.

So there are at least two places for the will to enter the picture when it comes to guessing. First, in settling whether one is to be agnostic or opinionated; and second, supposing one is to be opinionated, in breaking ties.

Note, however, that agnosticism can come in degrees. Consider now a horse race instead of a coin flip. Suppose there are four entrants, A, B, C, and D, whose respective chances of winning you know to be 40%, 30%, 20%, and 10%. Suppose you're asked what your guess is as to the outcome of the race (i.e., 'Who do you think will win?'). How might you answer? To my ears, any of the following replies would be entirely standard:²⁰

- A. ✓
- A or B. ✓
- Not D. ✓
- Either A, B, C, or D / Some horse or other. ✓

If you opt for the first answer, you are fully opinionated; if opt for the last, you are fully agnostic. By opting for one of the middle two, you are neither fully opinionated nor fully agnostic, but a mix of both. And which of these is your guess seems to be a matter of how you make up your mind. Once again, nothing about your estimation of the underlying probabilities forces your hand any which way. So it's up to you to decide not just *whether* you're opinionated on the answer to the question at hand, but *how* opinionated you are.

¹⁸ Calling this state a "guess" is a bit misleading, since you are sure that the coin will land either heads or tails, and we don't normally call things we are sure of *guesses*. But nothing of substance hangs on this: it's perfectly fine to answer in the affirmative if asked whether one *thinks* the coin will land either heads or tails (and in the negative if asked whether there is some particular way one thinks the coin will land). And for my purposes that's enough to count as having the relevant guess.

¹⁹ Things are slightly more subtle here, since one might start off having no particular guess, but then be unwilling or otherwise incapable of adopting a particular guess. In that sort of case, it's not clear that we'd want to describe one's agnosticism as having been arrived at will. Even then, what matters is that we *can* sometimes actively decide to be agnostic, much in the same way we can actively decide to remain still. And I see little reason for suspicion there.

²⁰ Note that on the intended interpretation of these answers, the proposition expressed by (e.g.) 'A or B' would be more exactly expressed by something like 'I don't have a view on which particular horse will win, but I do think it will be either A or B'. But for brevity's sake I will stick with the simpler answers in what follows. See Holguín (2022), Dorst & Mandelkern (2023), and Skipper (2023) for further discussion of the patterns here.

From here on out I will be taking for granted that we can and often do guess at will, both in breaking ties and in determining our degree of agnosticism. Of course, we can imagine situations in which one can't help but have some particular guess as to the outcome of a coin toss, or in which one simply "finds oneself" guessing that the winner of the horse race will be any horse other than D. Likewise, if we suppose (as I do) that (i) our states of surety are not determined at will, and that (ii) being sure of something entails having it as one's guess (i.e., that if you're sure that p , you think that p), then anytime you're asked to guess the answer to a question whose answer were sure of, your guess to that question will not be determined at will. But as before, this is all perfectly consistent with the claim that guesses can be and often are determined at will.

The previous section outlined a number of paradigm examples of voluntary acts and states. But I believe guessing provides the best model of the kind of willing I believe is involved in desire. In saying that we can and often do want at will, I am happy to limit myself to the claim that we can and often do form our desires in a manner that's as voluntary as the manner in which we form our guesses.

Dialectically, this puts constraints on the kinds of objections that can be raised against bouletic voluntarism. If an argument against the possibility of desiring at will can be transposed straightforwardly into an argument against the possibility of guessing at will, then we should not make much of that argument. To give a toy example: one might object that certain putative cases of wanting at will are *not* cases in which the relevant desires are formed voluntarily, but are instead cases in which the desires are the result of some kind of deterministic sub-personal process that (for whatever reason) is often accompanied by the distinctive phenomenology of choice. But absent reason to think that the same diagnosis cannot be applied to the "choices" underlying our guesses, we should be unmoved by this sort of objection. What we would have is not an objection to the claim that we desire at will, but a partial account of what *it is* to desire something at will.

I've gone through this detour on guessing not just because guessing helps further clarify the intended interpretation of 'at will', or because, dialectically, it serves as a catch-all for certain kinds of objections to bouletic voluntarism. I've gone through it because—as I soon hope to show—there is a deep structural parallel between guessing and desiring.

Just as our guesses fail to supervene on how likely we find things (our credences), so too do our desires fail to supervene on how desirable we find things (roughly: our subjective values). Two agents with identical credence functions can guess differently, since they can choose to break ties in different ways, or to adopt different levels of agnosticism. And two agents with identical "desirability functions" can have different desires, since they can choose to break evaluative ties in different ways, or to adopt different levels of *asceticism* (i.e., indifference to evaluative difference). It is in these respects that we can and often do want at will.

The next few sections of the paper will develop these ideas in a few steps. I'll start by unpacking the relevant notion of desirability (§5). Then I'll argue in favor of bouletic tie-breaking (§6)—the claim that our desires can discriminate between options we take to be equally desirable—and asceticism (§7)—the claim that our desires can fail to discriminate between options we take to be *unequally* desirable. As we'll see in §8, if these claims are right, then the re-

relationship between desire and desirability is considerably more complex than is suggested by standard theories of desire. I'll then turn to offering my own preferred account of the relationship (§9), which is in terms of a principle I call *cogency*—the bouletic analog of a principle that has been argued to govern the relationship between guessing and credence (Holguín 2022, Dorst & Mandelkern 2023). Only then in §10 will we finally see the direct argument for bouletic voluntarism.

5 Desirability

I am going to help myself to the assumption that, in addition to the facts about which propositions agents desire, there are facts about *how desirable* those propositions are for those agents.

Desirability, as I will understand it, is the kind of value relevant to desire—whatever that kind of value may be. I will assume that desirability is agent-relative and that it is connected to the relevant agent's interests, values, priorities, attractions, reasons, beliefs, and so on.²¹ I will assume, for example, that for typical human beings (*ceteris paribus*, etc., etc.): having more money is more desirable than having less; experiencing pleasure is more desirable than experiencing pain; an 80% chance of \$10 is more desirable than a 90% chance of \$5; and any two prospects that are in every relevant respect qualitatively identical are thereby equally desirable (such as an even-money bet on a fair coin landing heads versus an even-money bet on a fair coin landing tails).

There is a rich tradition of theorizing about desire in terms of desirability (or at least something akin to it). Following Phillips-Brown (2021), we can distinguish between two broad classes of theories of desire: “what’s best” theories and “what’s good enough” theories. Since we’ll be returning to both kinds of theories of desire in some detail in §8, for now I’ll describe them imprecisely (and abstract away from sectarian disputes).

According to the “what’s best” theory, to desire p is for p to be the “best” alternative in some relevant alternative set Q .²² What makes a proposition “best” (relative to Q)? Well, scoring highest on some relevant dimension of value. But what’s the relevant dimension of value? The answer to that question depends on which “what’s best” theorist you ask. Some take it as a theoretical primitive, some understand it in terms of the notion of (expected) utility familiar from Bayesian decision theory, and some understand it in terms of the agent’s preferences. But in common to all versions of the theory is the existence of a dimension of value on which prospects can be compared, with the unifying idea being that one desires whichever proposition scores highest on this dimension. I’m choosing to call the relevant kind of value ‘desirability’.

Likewise for the “what’s good enough” theory. On this view, to desire p is simply for p to be “sufficiently good”.²³ What makes a proposition sufficiently good? Scoring high enough on

²¹ Though I will not be assuming that facts about desirability are necessarily analyzable in terms of these sorts of facts. See Schroeder (2020) for a helpful overview of the different conceptions of (what I am calling) desirability that have been explored in the philosophical literature on desire.

²² See, e.g., Lewis (1974), Davis (1984), Stalnaker (1984), Pettit (1991, 2002), Heim (1992), Dreier (1996), von Fintel (1999), Levinson (2003), Condoravdi & Lauer (2016), Jerzak (2019). Note that for some “what’s best” theorists, Q is always the two-cell alternative set $\{p, \neg p\}$; for others, its value is sensitive to context and can sometimes include a number of different mutually exclusive propositions ($\{p, q, r, \dots\}$).

²³ See, e.g., Bradley (1999), van Rooij (1999), Lassiter (2011), Blumberg & Hawthorne (2021), and Phillips-Brown (2021, 2025).

some relevant dimension of value. Here too different theorists have different conceptions of the value in question. But they are united in theorizing in terms of such a value, which again I'm choosing to call 'desirability'.

Though I will be following the "what's best" and "what's good enough" theorists in invoking a notion of desirability and assigning it an important role in the theory of desire, I'm going to be defending a rather different conception of the relationship between these two notions. In particular, both the "what's best" and "what's good enough" theories take the facts about desirability to settle the facts about desire: knowing how desirable an agent finds certain propositions is enough to know which of those propositions the agent wants. I will deny this. And for reasons that will emerge later, this will require a *somewhat* freestanding conception of desirability.²⁴ In particular, it will require a conception of desirability on which we can regard questions like 'Can our desires discriminate between options we know to be equally desirable?' as non-trivial, and on which we can be reasonably confident in claims about how desirable certain propositions are for an agent even when we're comparatively less confident in claims about what that agent in fact desires. For the most part I will simply be taking for granted the intelligibility of these questions, but on occasion I will briefly engage with reasons for skepticism.

Officially, then, I'll be treating the notion of desirability as a theoretical primitive.²⁵ But I think we have a reasonably strong pre-theoretic understanding of the notion—enough to make plausible the kinds of judgments to which I've just alluded. I suspect, for example, that we can and often do express facts about (comparative) desirability using ordinary language expressions like 'better', 'more valuable', 'more appealing', and so on.²⁶ Indeed, I will often use judgments about sentences that contain these expressions in service of arguments about the relationship between desire and desirability. But I want to signal at the outset that these judgments need to be treated at arm's length. The intended interpretation of 'desirable' is *not* meant to be fully anchored to judgments about claims involving expressions like 'better', 'more valuable', and 'more appealing', since these expressions can easily be interpreted so that their connections to desire are tenuous at best. As I intend to use 'desirable', it could well be that 'Smoking is more desirable to *S* than quitting' is true on the intended interpretation even though 'It is better for *S* to quit smoking' is false on just about every natural interpretation. The facts about *desirability* are more intimately tied up to the facts about what an agent desires than to the facts about what seems good from their perspective.

This is all to say that the notion of *desirability* is a quasi-theoretical one, like the notion of *similarity* in the theory of counterfactuals (Stalnaker 1968, Lewis 1973), as well as the notions of *safety* and *normality* in the theory of knowledge (Williamson 2000, Goodman & Salow 2023, Carter & Hawthorne forthcoming). We can have a partial grasp of desirability independently of its role in the theory of desire. But we should not expect the two notions to be fully separable.

One final point. So far I've been talking as if there is such a thing as *the* desirability ordering associated with a given agent. But really I take there to be a *family* of such orderings, with different resolutions of the context-sensitivity of 'want' corresponding to variations

²⁴ I will be assuming a kind of "non-constructive realism" about desirability, to use the terminology of Dreier (1996) and Buchak (2013, ch. 1).

²⁵ We'll return to the question of its grounds in some detail in §12.

²⁶ Likewise 'desirable'. However, the ordinary notion of 'desirable' tends to mean something closer to 'is a worthy object of desire'. My semi-technical notion of 'desirable' is not meant to carry with it any normative implications.

in which of the members of this family are operative in context. Some of these desirability orderings—presumably most—are sensitive to the agent’s beliefs and evidence, but others may not be (cf. [Jerzak 2019](#)). Likewise, some of these orderings care about means (‘I want to wake up early so I can study’), while others may only care about certain kinds of ends (‘Of course I don’t want to wake up early; I’m just doing it so I can study’). In any event, as emphasized in §2, my intention is to ignore the context-sensitivity of ‘want’ when convenient. So in what follows I’ll mostly be talking as if each agent is associated with a unique desirability ordering.

With that, we may now begin working our way toward bouletic voluntarism.

6 Tie-breaking

This section will argue for the thesis that our desires (and preferences) can discriminate between options we know to be equally desirable.²⁷ That is:

TIE-BREAKING

It can be and often is the case that: (i) *S* knows that *p* and *q* are equally desirable, (ii) *S* wants *p*, but (iii) it is not the case that *S* wants *q*.²⁸

A paradigm example of choice in the face of equidesirability is the predicament of *Buridan’s ass*, starving and stuck between two equally appealing bales of hay. (I will call cases like it ‘Buridan cases’.) Most all of the existing philosophical literature on Buridan cases focuses on their implications for theories of (rational) *action*.²⁹ For they raise the question of whether it is possible to (rationally) pursue a course of action—say to go left rather than right—without believing that it is best or having most reason to pursue it.³⁰ This is a topic we will return to near the end of the paper. For now our focus will be on the relevance of Buridan cases to theories of desire—a topic that has been comparatively neglected.³¹

The idea that we can have discriminating desires (i.e., preferences) in Buridan cases has a fair amount of support.³² What is less widely supported is the idea that these discriminating desires can be formed *at will*. As these cases are typically described, the agents in question simply find themselves preferring one of the options (when they have any preference at all), and then act on the basis of that preference. Though eventually I’ll be arguing that agents who have desires in Buridan cases can and often do form those desires at will (§10), for now we can

²⁷ Without the qualifier about the options being *known* to be equally desirable, this thesis would be too easy to argue for, at least on certain ways of understanding ‘desirable’ (like those where the desirability facts are determined by the “objective” value of the options in question).

²⁸ *Mutatis mutandis* for preference: it can be and often is the case that: (i) *S* knows that *p* and *q* are equally desirable and (ii) *S* prefers *p* to *q*.

²⁹ The literature on this topic is old and expansive—see, e.g., [Al-Ghazali \(1095\)](#), [Rescher \(1960\)](#), [Ullmann-Margalit & Morgenbesser \(1977\)](#), [Mele \(1991\)](#), [Weintraub \(2012\)](#), [Chislenko \(2016\)](#).

³⁰ Some theorists, e.g., [Rescher \(1960\)](#), describe Buridan cases as those of action “in the absence of preference”. However, I doubt whether this is a suitably general characterization of the core phenomenon, since I will be arguing that agents can and do form preferences in the relevant cases. But all this is compatible with there being residual questions about action in cases where agents *don’t* form these preferences, or, when they do, of whether action grounded in them can be justified.

³¹ Judging by [Rescher’s \(1960, p. 159\)](#)’s translation, [Bayle \(1697\)](#) is a relevant exception.

³² For assent, see, e.g., [Mele \(1991\)](#), [Scanlon \(1998, p. 48\)](#), [Raz \(1999, p. 62\)](#), [Audi \(2001, p. 130\)](#), [Chang \(2004, 2009\)](#), [Tanyi \(2011\)](#).

set aside this matter. My goal here is just to establish the claim that we can form the relevant desires (whether or not at will).

So, here's a motivating example. Suppose you face two transparent boxes respectively labeled 'Left' and 'Right'. You can see that each contains a single \$20 bill. You are equidistant from both boxes, which, with the exception of their labels, are plainly qualitatively indistinguishable. In a moment you will be given the opportunity to open one of the two boxes and keep the \$20 bill you find inside. But first you must answer a question: Which particular box do you want to open?

To my ears, you would do perfectly well answering in any of the following ways:

- Left.
- Right.
- Either Left or Right / I don't care which.

Mutatis mutandis for preference: I detect no meaningful difference in the range of acceptable answers to a question like 'Which of the two boxes would you prefer to open?'. You might say 'Left', you might say 'Right', or you might simply deny that you have a preference between them.

Suppose you say 'Left'. Then it seems to follow that you want to open Left in particular (and that you prefer opening Left to Right). So you want Left in particular (and prefer Left to Right), despite the fact that Left and Right are equally desirable to you. Hence TIE-BREAKING.

Note that a similar argument can be run in the third-person. Imagine you're observing someone else in this situation. You know they're certain the two boxes are exactly alike in all relevant respects, and that they intend to be fully cooperative and sincere in everything they say. Do you know how they'll answer the question of which box they want (prefer) to open? I don't see how you could. For all you know they'll say 'Left', for all you know they'll say 'Right', and for all you know they'll say something like 'I don't have a preference' or 'Neither box is such that I want to open it in particular; I just want one of them'. Plausibly, if you knew what they wanted, you'd know how they would answer (again: you know they intend to be cooperative and sincere). But you don't know how they'll answer. So it must be that you don't know what they want until they tell you. So knowing that the options are equally desirable to them isn't enough to settle what they want, since they might well have a discriminating desire.

The particular Buridan case I've chosen to focus on is fairly artificial, but such cases abound in ordinary life. Most people seem to have no problem at all reporting discriminating desires when approaching forks in the road, selecting between cans of soup at the grocery store, picking a card in a magician's trick, or taking a side on an even-money bet. In typical cases, these preferences exist not because there is a perceived difference in the quality of the options, but because when the options are equally good, our desires are free to discriminate between them. This is much in the same way that when two competing answers to a question are equiprobable, our guesses are free to discriminate between them. The facts about probability are not always enough to settle the facts about our guesses. Likewise, the facts about desirability are not always enough to settle the facts about our desires.

That's the full positive case for TIE-BREAKING. Again, I want to emphasize that the claim is not that we *invariably* have discriminating desires in Buridan cases. It's just that it's not atypical for us to do so. I'll now consider and respond to some objections to these ideas.

Objection 1: Loose speech (broadly construed)

I've chosen to frame the argument from Buridan cases in terms of certain linguistic judgments. I've argued that when asked what we want in a Buridan case, we are free to say (e.g.) 'Left' rather than 'Right' (though we can also say 'I don't care'), and that this shows that we can form discriminating desires in the face of evaluative indifference. But one might think that talk of desire is one thing, genuine desire another. Perhaps all my argument really shows is that we sometimes speak *as if* we had discriminating desires in the face of evaluative indifference, rather than that we actually do have such desires.

To be clear, I see no reason to think that people are simply deluded about what they want in Buridan cases. So for this response to undermine the inference from language to mind, it needs to be that in saying (e.g.) 'I want/prefer Left', one invariably speaks either insincerely or non-literally. Methodologically speaking, I think there should be a strong presumption against this treatment of the judgments, since the practice of announcing preferences in Buridan cases bears none of the hallmarks of insincere or non-literal speech. But these general qualms aside, I will briefly explore two implementations of this style of response and explain why I find them wanting on their own terms.

The first version of the "loose speech" response is centered on the observation that wh-questions like 'Which *F* *G*s?' typically *presuppose* that there is an *F* that is *G*. So a question like 'Which box do you want/prefer?' will typically presuppose that there is a box such that the respondent wants/prefers that box. Thus, in being asked this question, one is under pragmatic pressure to report oneself as having a preference for a certain box, even if in fact one is indifferent between them. So in saying 'Left', one could easily be speaking insincerely, since whether or not one *really* prefers Left to Right, one has reason to say one does, at least for the sake of being cooperative.³³

I think there are a number of reasons for skepticism about this account of desire reports in Buridan cases. Here I'll mention three.

First, we can run the thought experiment with a question other than 'Which box do you want/prefer?', say: 'Which box, *if any*, do you want/prefer?', 'Do you want Left over Right?', or 'Do you have a preference for Left, a preference for Right, or no preference between them?'. These latter questions do not presuppose that there is a particular box you want, and yet the pattern of acceptable answers remains the same.

Second, suppose instead of asking 'Which box do you want/prefer?', I instead ask 'Which box do you need?' or 'Which box are you dying to have?'. To my ears, the only non-strange answer to these questions (at least in ordinary circumstances) is some version of 'Neither in particular'. But the coercive mechanisms are as active here as they are in the questions about what we want/prefer. The most natural explanation of this contrast is that the facts about our desires/preferences can discriminate between options in Buridan cases, while the facts about our needs or what we're dying to have cannot.

³³ Note that the choice about what to say here—i.e., uttering 'Left' versus uttering 'Right'—is itself a Buridan case (at least supposing one doesn't actually have a preference between Left and Right). Does this create a revenge problem for proponents of this account of the judgments? Not necessarily. If it is possible to intentionally ϕ rather than ψ without preferring ϕ -ing to ψ -ing, then it needn't follow from the fact that you have chosen to say 'Left' that you *prefer* to say 'Left'. We will return to the question of the relationship between intentional action and desire in §14.

Third, the focus on question/answer pairs is inessential, since the relevant mental states can be formed *in foro interno*. I think this is intuitively clear in its own right: I myself have no difficulty imagining Buridan cases in which I form arbitrary desires/preferences, whether or not anyone has prompted me to express them. But it is also suggested by the third-personal considerations raised above: if asked whether someone in the midst of a Buridan case has a preference between the two options, ‘I don’t know’ will often be a perfectly good answer, while ‘No’ will be presumptive, even in situations in which you are as certain as can be that the subject takes themselves to be in a Buridan case. So I don’t see concerns about the pragmatics of questions as getting to the heart of the matter.

A different version of the “loose speech” response claims not that desire reports in Buridan cases are typically insincere (because elicited by coercion), but that they involve a special—and perhaps non-literal—interpretation of ‘want’ and its cognates. In particular, one could argue that in saying ‘I want/prefer Left’, what one is invariably reporting is an *intention* to take Left. If this is right, then the relevant intuitions are not evidence in favor of the fact that our desires (in general) can break ties, but evidence in favor of the far less controversial thesis that our intentions can.

For the sake of argument I will assume that there is a general practice of sometimes interpreting ‘want’ so that the inference from ‘S intends *p*’ to ‘S wants *p*’ appears valid.³⁴ Even still, there are at least two reasons we should reject this as a general diagnosis of the judgments motivating TIE-BREAKING.

The first issue with this diagnosis is that it would suggest that it is impossible to intentionally hide one’s desires/preferences in Buridan cases. In particular, if the only true interpretation of ‘I want Left’ is one on which it expresses something tantamount to the proposition that I intend to open Left (rather than Right), then the only way I could be deceptive in saying ‘I want Left’ is if in fact I don’t intend to open Left (say because I intend to open Right). But it seems perfectly coherent that I might prefer to open Right (rather than Left), but even more strongly prefer to deceive you about my underlying preference. In such circumstances, it would be natural to intend to open Left—and then announce and act on that very intention—as a means of deceiving you about my underlying preference for opening Right. But this kind of deception would be impossible if Buridan desires were really just Buridan intentions.³⁵

However, there is a more serious issue with the intention diagnosis, which is that it is perfectly felicitous to express tie-breaking desires/preferences about propositions that could not be the objects of one’s intentions. For example: you can observe someone else, say Jones, in the midst of a Buridan case and say ‘I want Jones to open Left’ (even when you know full well that it otherwise makes no difference to you, or Jones, or anyone else, which option he takes). Similarly, if you’re told that one of the boxes was pre-selected for you to open and are asked which you want/prefer

³⁴ Cf. Nagel (1970, p. 29), who writes ‘...whatever may be the motivation for someone’s intentional pursuit of a goal, it becomes in virtue of his pursuit *ipso facto* appropriate to ascribe to him a desire for that goal’. See also Condoravdi & Lauer (2016).

³⁵ Granted, if my aim is to get you to falsely believe that I want to open Left (when in fact I want to open Right), there will plausibly be a sense of ‘want’ in which ‘I want to open Left’ expresses a truth, since opening Left rather than Right is what gets me what I want: namely, to deceive you about what I want. But this kind of non-uniformity in the interpretation of ‘want’ is something we should all be used to: witness the natural (albeit non-uniform) readings of speeches like: ‘Why do I want to buy the Leibniz book rather than the Kant book? Because it’s the best way to hide the fact that what I really want is to buy the Kant book rather than the Leibniz book’.

it to have been, ‘Left’ remains fine. Or suppose you’re in a hand of poker waiting for the final card to be dealt. You know you’ll win the hand if and only if one of the four remaining aces is turned over. If asked, ‘Which card do you want to be dealt here?’, an answer along the lines of ‘The ace of spades’ is perfectly standard. As anyone who has watched a bit of tournament poker will know, it is quite common for players to express a desire to see a particular card in “all-in” situations like this, even when they know full well that the card they name is one of many that will result in their being awarded the pot. So even if there is a sense of ‘want’ on which ‘S wants *p*’ is essentially synonymous with ‘S intends *p*’, cases like these show that the truth of TIE-BREAKING is not limited to such interpretations.

Objection 2: No true Buridan case

I’ve just explained why I think that someone who says ‘I want Left’ in the two box case expresses a genuine discriminating desire for Left over Right. But even supposing I’m right about this, there remains a path for resisting the argument for TIE-BREAKING, since one could always insist that such a case isn’t actually one in which the options under consideration are equidesirable. Perhaps it’s just a pseudo-Buridan case.

So, is there a way of proving that a choice between two equidistant boxes, each of which contains nothing more than a \$20 bill, is a true Buridan case? Well, here the situation is somewhat delicate, since what’s now under dispute is a question about the desirability of some hypothetical options for a hypothetical agent, where the relevant notion of ‘desirability’ is quasi-theoretical and so less than fully amenable to direct intuitive probing. Still, I think there is a strong case to be made that the kinds of examples I’ve been considering constitute genuine Buridan cases, and that the best arguments for thinking otherwise are ultimately unpersuasive.

There is a tradition of denying that an agent could ever be in a situation where they face two or more equally *choice-worthy* options—a kind of skepticism about the very possibility of genuine Buridan cases, at least for decision-making. [Leibniz \(1952\)](#) is famous for holding this view.³⁶ Now, I don’t find it plausible that *choice-worthiness* is one and the same thing as *desirability*. This is because propositions that are not and perhaps *could not be* the objects of choice can nonetheless vary in their desirability. But I’m happy to suppose that the Leibnizian view, if sound, can be generalized to show the impossibility of cases of equidesirability.

However, I follow [Ullmann-Margalit & Morgenbesser \(1977\)](#), [Bratman \(1987, 1999\)](#), and [Chislenko \(2016\)](#) (among others) in rejecting Leibniz-style skepticism about Buridan cases. Take the case of the two boxes. If the Leibnizian view is right, then one of the boxes is more choice-worthy (desirable) than the other. Let’s suppose it’s the Left one. The question is what could make this so.³⁷ Is it the serial numbers on the \$20 bill inside? Presumably not. But even if they did make a relevant difference, couldn’t we simply account for the symmetry-breaking effects of the serial numbers (and whatever else) by adding pennies—or, better, increasingly marginal lotteries for pennies—to whichever box is inferior, until eventually we rebalance the scales and arrive at a proper Buridan case? If not, then by what magic do the scales always manage to

³⁶ See the epigraph. [Montaigne \(1958\)](#) is another influential proponent of this view.

³⁷ Here I’m presupposing that it’s not just a brute fact that Left is more choice-worthy/desirable than Right. Leibniz, of course, would accept this presupposition. But there is in principle room to reject it.

resist this sort of rebalancing? As I see it, to deny the possibility of Buridan cases *tout court* is to take the path of darkness.

The better path, in my view, involves a more circumscribed skepticism about Buridan cases. In particular, one could try positing a connection between desire and desirability that leaves unscathed the possibility of Buridan cases, but eliminates the possibility of tie-breaking. What might such a connection look like? Well, the natural idea would be to insist that desire and desirability are connected in such a way that *being equally desirable* metaphysically necessitates *being indifferent between*. For then Buridan cases would be possible, but discriminating desires in Buridan cases would not, since the proposition that *S* has a discriminating desire for Left over Right (say) would entail that Left is more desirable to *S* than Right. In short: the mark of a true Buridan case is that you don't and in fact *can't* have a discriminating desire.

One way of developing this response involves a subtle variation on the Leibnizian perspective. In particular, one could insist that it's just a necessary truth that *if S* prefers Left to Right, then Left has some *quality* in virtue of which it is more desirable to *S* than Right. Maybe it's the serial numbers. Maybe *S* just happens to have a brute fondness for the sinistral over the dextral. Who knows. In any case, the guiding thought is that there must be *some* such tie-breaker, since (so the thought goes) *S*'s desires can never be arbitrary by the lights of *S*'s conception of the value of things (desire must have some reason or principle). And unlike the crude Leibnizian position—which denies the very possibility of Buridan cases—this position does not bring with it the same sort of metaphysical baggage. The scales can in principle be balanced: by adding pennies (or lotteries for pennies) to Right, one can eventually get the two options to be equally desirable; it's just that when this finally happens, *S* will no longer have a preference. This perspective thus doesn't suffer from the main shortcoming of the cruder forms of skepticism about Buridan cases.

Unfortunately, this more sophisticated perspective is problematic too. The claim that one's desires cannot be formed arbitrarily—that desire must have some reason or principle—is not an analytic truth.³⁸ It's a substantive philosophical thesis, and one that we should in principle be willing to give up in the face of contravening evidence. And the basic intuitive judgments about (putative) Buridan cases strike me as strong contravening evidence indeed. If not cases like these, then what could possibly shake one's commitment to the principle that discriminating desires cannot be formed arbitrarily? Yes, we can imagine agents that prefer Left to Right not arbitrarily, but because of their proclivity for certain serial numbers or sinistrality or what have you. But I see nothing whatsoever to suggest that we'll always be able to find such Leibnizian tie-breakers, if only we squint hard enough.

The analogy to guessing is helpful here. The claim that one's guesses cannot be formed arbitrarily (by the lights of one's credence function) is not an analytic truth. It's a substantive philosophical thesis, and one that we should be willing to give up in the face of contravening evidence. And the fact that we can and often do guess the outcome of fair coin tosses is strong contravening evidence indeed. Yes, we can imagine agents who guess Heads rather than Tails not arbitrarily, but because they have some general superstition to the effect that "fair" coins land heads more often than tails. But I see nothing whatsoever to suggest we'll always be able

³⁸ At least not in its current form. We'll consider such variations in a moment.

to find such Leibnizian tie-breakers, if only we squint hard enough.

There are also some direct problems with the sophisticated Leibnizian perspective. Suppose *S* prefers Left to Right. By hypothesis, then, *S* finds Left more desirable than Right. But just how much more desirable? Well, suppose that although *S* prefers Left to Right when each contains \$20, *S*'s preferences would flip were an extra dollar added to Right. If *S* is like that, then it seems to follow that the original desirability gap (for *S*) between Left and Right must have been less than the desirability of a dollar, since a dollar was enough to tip the scales in the other direction. But surely the same could well be true if instead of a dollar we used a penny, or a coin-flip for a penny, or a coin-flip for a coin-flip for a penny: in each of these cases we can imagine that *S* would prefer the (ever more marginally) sweetened Right to Left, further bounding the original desirability gap between Left and Right. But if Left really is more desirable (to *S*) than (unsweetened) Right, it seems there would have to be *some* sufficiently marginal lottery such that: even conditional on Right being sweetened by that lottery, *S* would continue to prefer Left. But as far as I can tell we can easily imagine otherwise. We can easily imagine *S*'s preference for Left being maximally fickle: *any* amount of sweetening induces a preference for Right, no matter how small. Supposing *S* is like this, it would look to follow that the size of the (original) desirability gap between Left and Right is like the size of the “gap” between $\sum_{n=1}^N \frac{9}{10^n}$ and 1 as *N* approaches infinity—zero. The sophisticated Leibnizian thus appears to be committed to the idea that anyone who prefers Left to Right will in principle continue to prefer Left to Right in the face of at least some amount of Right-sweetening. But this too seems to me an implausible article of faith.

A different strategy for resisting the argument for TIE-BREAKING involves positing a *conceptual* connection between desire and desirability—one that essentially rules out the possibility of TIE-BREAKING by fiat. Perhaps it's just part of what ‘desire’ and ‘desirable’ *mean* that if *S* desires *p* but doesn't desire *q*, then *S* finds *p* more desirable than *q*. This claim would be naturally supported by the idea that *being desirable* is best understood along the lines of *being conducive to the satisfaction of one's desires*, or something to that effect. For then the idea would be that if *S* prefers Left to Right, then *S* thereby finds Left more desirable than Right, since Left satisfies strictly more of *S*'s desires than Right.

My response to this treatment of Buridan cases is more dialectical than substantive. But supposing the view is right, then yes: the fact that *S* has a discriminating desire for Left no longer implies TIE-BREAKING—at least not as stated. And that's because, by hypothesis, the fact that *S* desires Left entails that Left is more desirable to *S* than Right. But what's supposed to explain this comparative desirability fact? Supposing the proponent of the view isn't smuggling in some Leibnizian tie-breaker—(like that Left is more desirable because the \$20 bill it contains has a nicer serial number)—then they're going to be on the hook for thinking that two options can be unequally desirable even when they are equally valuable by the lights of every other reasonable conception of value.³⁹ But once they've agreed to *that*, then (a) they're agreeing to

³⁹ Granted, the inference to a Leibnizian tie-breaker is more palatable in the third-person version of the case, since knowing that someone else wants Left rather than Right does seem to give one grounds to suspect that the person sees some relevant difference in the quality of the two boxes after all. (Maybe the observed person cares about serial numbers?) Still, my sense is that even after observing the person say ‘I want Left’, it would be hasty to *conclude* that the person takes the two options to be unequally good. Here the analogy to guessing is once again helpful. Learning that someone has guessed *p* in answer to the question ‘*p* or $\neg p$?’ provides *some* evidence that they think *p* is more

a thesis that is awfully close to TIE-BREAKING and (b) they're on the path to defending bouletic voluntarism in all but name.

Put another way: desirability is the kind of value relevant to desire, and whatever that kind of value turns out to be, it's implausible that it would be sensitive to whatever differences there happen to be between Left and Right. For typical agents, the two boxes are plainly equally good, in all relevant senses of 'equally good'. If there is an exception to this generalization, it's in the sense of 'equally good' (should there be one) on which it expresses something like *being equally conducive to the satisfaction of one's desires*. But then the essential points go through all the same, because all the questions we might have asked about desire we can now ask about the property of being equally conducive to the satisfaction of one's desires. We'll have good reason to believe that this property fails to supervene on other kinds of value—since it allows for a kind of tie-breaking—and we'll have good reason to believe that the facts about desirability can be and often are a matter of our choosing. In effect, then, we'll just have shuffled some labels around.

I thus conclude that being in a Buridan case is no barrier to having a discriminating desire: we can and often do form preferences in the face of equally desirable options. I believe this has important and largely unrecognized implications for the theory of desire, especially in connection to issues of voluntarism. Before we get into any of that, however, I want to discuss another striking aspect of the relationship between desire and desirability.

7 Asceticism

As discussed in §4, there are two respects in which guessing fails to supervene on credence. One respect is that our guesses can and often do break credal ties, as when we're guessing the outcome of a fair coin toss. I've just argued that desire is analogous to guessing in this way, in the sense that our desires can break desirability ties in Buridan cases.

But there is a second respect in which guessing fails to supervene on credence, which concerns one's level of *opinionatedness*. It is possible for someone to know that p is more likely than $\neg p$, and yet fail to guess that p . Just because you know a coin is biased 60:40 in favor of heads doesn't mean that your guess as to the answer to the question 'How will the coin land when flipped?' must be heads. Again, you might be *agnostic*, willing only to guess that the coin will land some way or other.

I believe desire is analogous to guessing in this second respect as well. In particular, I believe that the inference from ' p is more desirable to S than $\neg p$ ' to ' S wants p '/' S prefers p to $\neg p$ ' is invalid—much in the same way that the inference from ' p is more likely to S than not' to ' S thinks that p ' is invalid.⁴⁰ This is because we can be *ascetics*: agents who fail to form discriminating desires between options known not to be equally desirable. We can sincerely and truthfully say 'I don't care whether p ' or 'It's not correct to say that I want p ' or 'I don't prefer p to $\neg p$ ' even when we know full well that things will go better for us in all possibly relevant respects conditional on p . In short, we can be indifferent to (known) difference.⁴¹

likely than $\neg p$. But clearly it does not provide enough evidence to license a full-blown conclusion to that effect, since it remains compatible with the evidence that the guesser is engaging in some arbitrary tie-breaking, like in a coin toss scenario.

⁴⁰ This is not to say that these inferences are unreliable. More on this issue in §14.

⁴¹ As in the case of TIE-BREAKING, the principle is stated in terms of *known* difference so as not to render it trivial (on

ASCETICISM

It can be and often is the case that: (i) S knows that p is more desirable than $\neg p$ and (ii) it is not the case that S wants p .⁴²

The idea that our desires can fail to discriminate between options that we know to be equally good is not new. It can be found in various places in some of the more recent literature on desire, and is in fact one of the central considerations cited by proponents of the “what’s good enough” theory of desire in support of their view (see §5 above and §8 below). To my knowledge, however, none of these theorists take the possibility of asceticism to suggest that desire fails to supervene on desirability, or to tell in favor of bouletic voluntarism. I’ll draw out these connections later. For now I’ll just make the case for ASCETICISM.

Suppose you learn that a wealthy benefactor is about to deposit \$1,000,000 in your bank account, no strings attached. Suppose you’re then asked whether you would like to sweeten the deal by having an additional \$1 deposited (again no strings attached). How might you reply?

Clearly it would be fine for you to respond ‘Yes, I want the free \$1’ or ‘I prefer having the extra dollar to not having it’. But are these the *only* reasonable kinds of responses?

I think not. For it also seems clearly fine to respond ‘I don’t care’ or ‘I don’t have a preference’. After all, it’s just \$1, and you’re already about to get a million! There doesn’t seem to be anything particularly strange about being indifferent to the difference between having \$1,000,000 and having \$1,000,001 (cf. [Slote 1989](#)).

Expressions of indifference of this sort are ubiquitous in ordinary life. Suppose you’re at the grocery store choosing between a can of soup that expires in a year plus a day versus one that expires in a year plus a week. ‘No preference’ is fine. Or suppose you have a choice between two glasses of water, one of which is room temperature and one of which is half a degree celsius cooler than room temperature. ‘No preference’ is fine. Or suppose you’re pre-selecting seats for the movie you’re seeing tonight, where the only two available are one that is immediately left of dead-center and one that is two left of dead-center. ‘No preference’ is fine. Moreover, in each of these cases the person who sincerely claims not to have a preference can do so while recognizing that the options before them are *not* equally good. A response along the lines of ‘Yes, that one is slightly better—but who cares?’ may not be mandatory, but it’s also not atypical.

Similar points go through in the third-person. Knowing that someone else knows they’d be better off with the extra dollar does not in general suffice for knowing whether they want/prefer the dollar. They might say they prefer it (perhaps it’s even quite likely they will); but they might not. Until they answer, we should recognize the genuine possibility that they are indifferent to the difference. Likewise for the other examples mentioned in the previous paragraph: unless you know how picky the person is (with respect to the kinds of goods at issue), it’s not clear how you could know whether they have a preference. Maybe they regard it small enough to be disregarded.

That’s the positive case for ASCETICISM. Again, I want to emphasize that the claim is not that human beings are *invariably* indifferent to (small) evaluative differences. It’s just that it’s not

certain understandings of ‘desirability’).

⁴² *Mutatis mutandis* for preference: it can be and often is the case that: (i) p is more desirable than $\neg p$ and (ii) it is not the case that S prefers p to $\neg p$.

atypical for some of us to be. I'll now consider and respond to some objections to these ideas.⁴³

Objection 1: Loose speech, broadly construed

Like the positive case for TIE-BREAKING, the positive case for ASCETICISM has been framed in terms of linguistic judgments: in situations where it seems clear that we're deciding between options that are unequally desirable, we nonetheless can and often do say that our desires do not discriminate between those options. But perhaps talk of indifference to difference is one thing, genuine indifference to difference another. Are there reasons to regard the relevant linguistic judgments with suspicion?

As a first pass, one might wonder whether the relevant expressions of indifference are simply a byproduct of socio-pragmatic forces. Maybe it's just *uncouth* to report oneself as having a preference for the extra \$1 or the ever-so-slightly cooler glass of water, even though—strictly, literally, speaking—those are one's true preferences. (Who wants to seem so picky?)

My sense is that considerations of propriety are a genuine source of noise here, but not a serious one. For starters, there's the fact that all the relevant judgments can be made *in foro interno* (just like with tie-breaking). It's also not clear how this sort of thing would help explain our apparent ignorance of whether others are indifferent to known differences.

But even waiving these concerns and staying at the level of first-personal speech, it's just not that difficult to control for socio-pragmatic factors. Just imagine that you are under a veil of anonymity and in a setting where everyone's interest is exclusively in knowing the truth about what people want ('Please just be honest: do you want/prefer an additional \$1?'). Even in this setting it is perfectly fine to deny having a desire/preference for the extra \$1.

Instead of appealing to considerations of propriety, then, one might try to explain away the relevant linguistic judgments as involving "non-standard" (or otherwise idiosyncratic) resolutions of the context-sensitivity of 'want' and its cognates. Maybe in the contexts where it's fine to say 'It's not fair to say I want the \$1' or 'I don't have a preference for the extra \$1', what one is really saying is what one would (in a "normal" context) express with something like: 'It's not fair to say I *strongly* desire the \$1' or 'I don't have a *significant* preference for the extra \$1'. And it's no surprise at all that ASCETICISM could be true when 'want' and 'prefers' are interpreted as

⁴³ There is one style of the argument for ASCETICISM that I will not be pursuing here, which relates to *anhedonia*. Consider Stocker's (1979, p. 744) description of the phenomenon:

Lack of this desire is commonplace. Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such "depressions" is that one sees all the good to be won or saved and one lacks the will, interest, desire, or strength.

It doesn't take too much work to transform these observations about anhedonia into an argument for ASCETICISM. However, I think the issues here are messy enough that it is better not to rely on this sort of argument. In particular, I find it hard to tell whether the anhedonic person's failure to desire (let alone pursue) "the good" shows that they are failing to prefer more desirable options to less desirable options. An alternative diagnosis is that their condition carries with it a certain misalignment between what they find desirable (sitting in bed, ruminating, etc.) and what they believe they ought to find desirable (seeing friends, being active, etc.). Alternatively, we might think that the anhedonic person simply doesn't find *anything* particularly desirable, with life as one big bland Buridan case, as it were. On either of these possibilities—neither of which I want to rule out in advance—anhedonia needn't begot asceticism.

expressing these more demanding bouletic relations. However, I see two related reasons to be suspicious of this account of the judgments.

First, for this reinterpretation strategy to be at all plausible, it better be that there are readily available interpretations of ‘want’ on which one would intuitively speak *falsely* in uttering ‘It’s not fair to say that I want the \$1’. But it’s not clear there are any such interpretations. The taxation case provides a useful contrast here. If you ask me whether I want to pay my taxes, an answer along the lines of ‘Yes and no’ is easy to understand, and can be elaborated upon in a way that makes it perfectly clear what’s going on (‘... Yes, because I don’t want to go to jail. No, because I’d rather spend the money on frivolities’). The same sort of reply is much harder to understand in the \$1 giveaway: ‘Do I want a free \$1. Well, yes and no. Yes, because I’m technically better off with the free \$1. No, because I’m not all that much better off with it’. This sort of speech sounds bizarre in a way that the taxation speech does not: to my ears, it sounds like the speaker is simply changing their mind about whether they care enough about a free dollar to want it. This contrast is mysterious on the assumption that the only true interpretations of the asceticism speeches are those on which ‘want’ expresses a kind of strengthened desire.

Second, this account of the judgments implies that knowing that (i) *S* doesn’t strongly desire *p* suffices for knowing that (ii) ‘*S* doesn’t desire *p*’ expresses a truth in certain contexts (namely those contexts in which declarations of asceticism are felicitous). So suppose I know that Jones and myself are alike in that we don’t strongly desire the free \$1 (maybe Jones has announced much). Does it follow, then, that in the contexts where I can felicitously say ‘It’s not fair to say that I want the free \$1’—contexts in which, by hypothesis, I’m expressing the fact that I don’t strongly desire the free \$1—I can just as well say ‘It’s not fair to say that Jones wants the free \$1’?

I don’t see why it would. To my ears, anyway, such a speech is plainly presumptuous, and its presumptive nature is brought out by the fact that, for all I know, Jones might easily say something like ‘Although I don’t have a strong preference for the free \$1, I do have a preference for it’. This suggests that ‘want’ cannot so easily be read as synonymous with ‘strongly desire’.

Perhaps, then, the opponent of ASCETICISM should just go in for a blunter error-theory. Given that it’s more desirable to have a free \$1 than not, you simply can’t speak truly in saying that you fail to desire the free \$1. Those who sincerely say things like ‘It’s not fair to say I want the free \$1’ are either confused about what they want or, perhaps more charitably, about what ‘want’ means.

My main complaint with the blunt error-theory is comparably blunt: it’s just hard to see what independent motivation there could be for it. Yes, on certain *technical* interpretations of expressions like ‘desire’ and ‘preference’—such as those explicitly anchored to explanatory theories of behavior—the relevant connections between desire/preference and desirability might be strong enough to make asceticism (so understood) impossible. But absent some premise to the effect that the relations picked out by ‘desire’ and ‘preference’ in certain theoretical settings are the same as those picked out by ‘desire’ and ‘preference’ in ordinary settings, this kind of response is a non-sequitur. Moreover, all the considerations we’ve been surveying in this (and the previous) section look like compelling evidence against this premise.

One way to see the problem is to consider the dialectic when applied to bouletic expressions

other than ‘desire’ and ‘preference’—say ‘need’ and ‘wish’. In the context of the \$1,000,000 giveaway, speeches like ‘I’d be better off with the extra \$1, so I need the extra dollar’ are terrible, while speeches like ‘I’d be better off with the extra \$1, so I wish I had the extra dollar’ are perhaps not terrible, but still a bit craven. In any event, it’s perfectly fine to say things like ‘I’d be better off with the extra \$1, but I wouldn’t go so far as to say I need it/wish I had it’. Nor is there any hint that one who does so must be making a mistake (whether about mind or meaning). An error-theory about *these* kinds of judgments looks beyond the pale. But once that’s acknowledged, it’s not clear what pre-theoretic reason we’d have to expect things to be so different with desire/preference. After all, we seem to talk as if it’s fine for one’s desires and preferences to fail to discriminate between options that are known not to be equally desirable. So why should we have much confidence in theories of desire—*ordinary* desire—that predict that such speeches are guaranteed to express falsehoods? Better, in my view, to accept that the connections between the ordinary notion of desire and the values to which it’s responsive are less direct than is suggested by our use of the terms ‘desire’ and ‘preference’ in certain theoretical settings.

Objection 2: Equidesirable after all?

I’ve just argued that someone who says ‘It’s not fair to say I want the free \$1’ in the giveaway case expresses a genuine lack of preference for a free \$1 over the status quo. But even supposing I’m right about this, there remains a path for resisting the argument for ASCETICISM, since one could always insist that such a case is *not* one in which the options under consideration ($\{\$0, \$1\}$) are unequally desirable. Perhaps anyone who doesn’t want the free \$1 must thereby regard it as contributing nothing desirable to the state of things.

Once again the quasi-theoretical nature of desirability makes it difficult to adjudicate questions like this directly. Still, I think there is a strong case to be made against the view that indifference necessitates equidesirability.

For one, the proponent of this view owes us an account of the felicity of speeches like: ‘Yes, I recognize I’d be better off with an extra \$1, but it’s not fair to say that I want it—the difference is so small!’. For taken at face value, their truth implies ASCETICISM—at least so long as the relevant interpretation of ‘better’ is one on which it implies the corresponding desirability-theoretic claim. So it better be that the true interpretations of these speeches exclusively occur in contexts where ‘\$1 is better than \$0’ expresses a truth while ‘\$1 is more desirable than \$0’ expresses a falsehood. Can this conjecture be sustained?

Well, as stressed in §5, it is not difficult to find interpretations of ‘better’ on which, for certain p and q , ‘ p is better than q ’ expresses a truth while ‘ p is more desirable than q ’ expresses a falsehood. Claims about the relative value of options can track notions of value whose connections to desire (and thus desirability) are anywhere from indirect to non-existent. However, it’s hard to see how this sort of gap could be present when p is the proposition that the agent gets a free dollar and q is the proposition that the agent gets nothing. I see no reasonable interpretation of ‘desirable’ on which, for some fairly typical person, having an extra free dollar could fail to be knowably more desirable than not having an extra free dollar. I invite those who disagree (maybe money corrupts the soul?) to reimagine the giveaway in terms of hedons, nanoseconds

in heaven, or what have you. So long as there is some kind of value to which an agent's desires are responsive, I conjecture that we will be able to find cases in which that agent, faced with two options that differ ever so slightly with respect to that value, could reasonably say something to the effect of 'I know they're not equally good, but they're close enough that I don't have a preference between them'. Granted, I don't know how to *prove* a claim like this. But I submit that its denial looks difficult to sustain.⁴⁴

Independently of these more speculative points, there are also direct reasons to worry about the inference from indifference to equidesirability, related to the non-transitivity of indifference. I'll unpack this worry in steps.

For indifference to be non-transitive is for there to be some propositions p, q, r and some agent S such that: possibly, S is indifferent between p and q and between q and r , but *not* between p and r . As it turns out, the claim that indifference is non-transitive is effectively implied by the claim that asceticism can come in degrees. One who is indifferent to small differences (say, $\pm \$1$) needn't thereby be indifferent to large differences (say, $\pm \$1,000$): the former involves a milder form of asceticism than the latter. And this kind of differential indifference to difference brings the non-transitivity of indifference in its wake.

For example: suppose you're asked, 'You're about to get a free $\$n$ or $\$(n+1)$ —which do you want it to be?'. I submit that you could speak truly in saying 'I don't care which', no matter the value of n . (It's just a dollar, after all.) Holding this pattern of answers fixed, suppose you're then asked, 'You're about to get a free $\$0$ or $\$1,000$ —which do you want it to be?'. I submit that you could speak truly in saying 'I want the $\$1,000$ '. Indeed, you could even issue a general statement to the effect of: 'If the difference is just $\$1$, I don't have a preference; but if it's $\$1,000$ I do'. Taking these speeches at face value—which is what we're doing here—we have that, for any n , you are indifferent between $\$(n+1)$ and $\$n$, but also that you prefer $\$1,000$ to $\$0$. So you are indifferent between $\$1,000$ and $\$999$, between $\$999$ and $\$998$, ..., and between $\$1$ and $\$0$; but you are *not* indifferent between $\$1,000$ and $\$0$. So you are a witness to the non-transitivity of indifference.⁴⁵

Why does the non-transitivity of indifference pose a problem for the claim that indifference entails equidesirability? Because plausibly *being equally desirable* is an equivalence relation, and thus transitive, and so plausibly if p and q are equally desirable, and q and r are equally desirable, then p and r are equally desirable too. So if equidesirability corresponds to an equivalence

⁴⁴ Unless the denial is sustained by fiat. One could always insist that desirability must be understood directly in terms of the agent's desires/preference, in such a way that, by stipulation, *being indifferent between* entails *being equally desirable*. Interpreted this way, ASCETICISM would be trivially false. But dialectically the situation would be essentially the same as it was in the case of TIE-BREAKING: opponents of ASCETICISM will have secured a pyrrhic victory, since they'd be on the path to thinking (a) that the facts about desirability do not supervene on the facts about other kinds of value and (b) that voluntarism about desirability is true.

⁴⁵ The idea that indifference can be non-transitive has a long history—for discussion see, e.g., Armstrong (1950), Luce (1956), Sen (1969), Schick (1969), Fishburn (1973), MacKay (1980), Lehrer & Wagner (1985), Schumm (1987), Quinn (1990), Anand (1993), Hansson (1993), Gendin (1996), Danielsson (1998), Aldred (2007), and Baumann (2022). Two things to note here, however. First, the relevant notion of 'indifference' in this literature is often tied to a quasi-technical conception of 'preference', operationalized in terms of dispositions to act or choose. It's thus not entirely obvious whether these arguments can be successfully adapted to show that *ordinary* indifference (i.e., an absence of *ordinary* preference) is non-transitive. Second, many of the existing arguments for the non-transitivity of "indifference" are motivated by considerations of *indiscriminability*, with the crucial premise being that indiscriminability between options (with respect to their relevant dimensions of value) suffices for indifference between these options. Since indiscriminability is non-transitive (Williamson 1990), it would follow that indifference is too. However I want to emphasize that this is not *my* argument for the non-transitivity of indifference, since my argument is grounded entirely in terms of the possibility of (moderate) asceticism in the face of *known* difference.

relation while indifference corresponds to a non-transitive relation, then indifference cannot entail equidesirability.

Why think that *being equally desirable* is an equivalence relation? Because desirability is the kind of value to which our desires are responsive, and whatever that kind of value is, it would be surprising if it allowed for equidesirability to fail to correspond to an equivalence relation. Again, though *being desirable* is not meant to be reducible in any straightforward way to *being* Φ , where Φ might be any of: *appealing*, *conducive to one's interests*, *pleasurable*, *good*, etc.—any such Φ is one for which *being equally* Φ is plausibly an equivalence relation. So if *being equally desirable* isn't an equivalence relation, then in virtue of what could this be so?

Now admittedly a rhetorical question like this can only carry so much force. But its force can be amplified by seeing how one might go about actually trying to answer it. For as far as I can tell, the most plausible conceptions of desirability on which *being equally desirable* fails to be an equivalence relation are either dialectically inert or independently problematic. I'll consider two such conceptions here.

First (and as before), one might think that 'desirability' has to be understood directly in terms of desire/preference, so that, as a conceptual matter, 'S is indifferent between p and q ' entails ' p and q are equally desirable to S'. On this interpretation of 'desirable', the non-transitivity of indifference simply entails the non-transitivity of equidesirability (and thus its failure to be an equivalence relation). But again, not much is accomplished in interpreting 'desirable' this way, since all the questions we've been asking about the relationship between desire and desirability can just be reinterpreted as questions about the relationship between desirability and the other kinds of values to which it's connected. We'll still have the relevant kinds of supervenience failures (now: desirability on other kinds of value; before: desire on desirability), and with them the essential materials for the argument for bouletic voluntarism.

Second, one might take the facts about desirability to be intimately connected to the facts about *discriminable* value. In particular, one might think that the inference from 'S cannot make any relevant discriminations between p and q ' to ' p and q are equally desirable (to S)' is valid, and thus that the non-transitivity of discriminability implies the non-transitivity of desirability (cf. footnote 45). This sort of view would have the advantage of making sense of how: (i) desirability has something important to do with Φ (where Φ is some combination of the kinds of values listed above); (ii) *being equally* Φ is an equivalence relation; but (iii) *being equally desirable* is not. The idea would simply be that it's only the discriminable differences in Φ -ness that correspond to differences in desirability.

Still, I don't think this response accomplishes much. For starters, even if we can be and often are *indifferent* between options whose value we cannot discriminate—and so even if the non-transitivity of indiscriminability lends itself to the non-transitivity of indifference—this does not imply that options that are indiscriminably valuable are thereby equally *desirable*. Yes, desirability is a kind of subjective value, and so facts about desirability tend to depend on the facts about what an agent is in a position to know (*inter alia*). But all this is perfectly compatible with the idea that there can be cases in which an agent isn't in a position to know of two unequally desirable options that they are in fact unequally desirable; being subjective is not the same thing

as being transparent.⁴⁶

More substantively though: even if we wanted to hard-wire the inference from ‘indiscriminable’ to ‘equidesirable’ into our account of desirability, there remains the fact that the value-relevant differences between \$0 and \$1, though small, are plainly discriminable. So even if *being equally desirable* is not in general transitive (and thus not an equivalence relation), it plausibly behaves transitively when restricted to certain kinds of prospects and values—like known quantities of money. And that kind of “local” transitivity is enough for present purposes.

Here, then, is the essence of the problem for the view that whenever one fails to prefer \$1 to \$0, it’s inevitably because one regards \$1 and \$0 as equally desirable. If \$1 and \$0 really were equally desirable, then, *ceteris paribus*, we’d expect that, for any arbitrary amount of money, the prospect of that amount plus \$1 is exactly as desirable as the prospect of that amount plus \$0.⁴⁷ But since (financial) asceticism can come in degrees, you might well prefer \$1,000 to \$0, despite failing to prefer \$1 to \$0. If you are such a person, then there is a least number $1 < n \leq 1,000$ such that you prefer \$ n to \$0, but fail to prefer \$ $n-1$ to \$0. So sweetening the \$1 with \$ $n-1$ dollars is enough to induce a preference (you’ll prefer the \$ n to the \$0), while sweetening the \$0 with \$ $n-1$ dollars is not (you’ll be indifferent between the \$ $n-1$ and the \$1). So despite the fact that you are indifferent between \$1 and \$0, we should deny that they are equally desirable to you, since they don’t respond the same way to equivalent sweetenings. Indifference to a free \$1 is a genuine form of asceticism.⁴⁸

I thus conclude that known difference is no barrier to indifference: we can and often do fail to form discriminating desires in the face of unequally desirable options. As with tie-breaking, I believe this has important and largely unrecognized implications for the theory of desire, especially in connection to issues of voluntarism. I now turn to exploring these implications in detail.

8 Against existing theories of desire

I’ll start by explaining why existing theories of desire are ill-suited to account for TIE-BREAKING and ASCETICISM. I will focus on two broad classes of theories of desire, both introduced briefly in §5 above: the “what’s best” and “what’s good enough” theories. I will argue they are alike in sharing the same principal shortcoming: they take facts about desirability to *settle* the facts about desire. As we’ll see in due course, the lesson of TIE-BREAKING and ASCETICISM is that the relationship between desire and desirability has to be considerably less direct than that.

⁴⁶ Here it is worth noting that the standard ways of modeling non-transitive indifference (see, e.g., Luce 1956) involve an underlying dimension of value that *does* discriminate between options (even though those options are indiscriminable to the agent in question).

⁴⁷ The ‘*ceteris paribus*’ clause is needed to control for cases where, e.g., it’s illegal to have exactly \$1,000,001 in one’s bank account (while any other amount is permitted).

⁴⁸ Notably this kind of argument is not meant to be compelling in cases where the relative (financial) value of the unsweetened options is indiscriminable. Suppose instead of \$1 and \$0, the choice was between fist-sized pile of gold dust A and fist-sized pile of gold dust B. Suppose A and B are indiscriminably large: you can’t tell by looking which is larger, or if in fact they’re exactly the same size. Suppose, as such, that you’re indifferent between them, but also that if one was clearly larger than the other, you’d have a preference for that one. Suppose, finally, that unbeknownst to you, B is in fact slightly larger than A (perhaps only by a speck or two). Then, plausibly, there is some amount of dust x such that adding x to B induces a preference for B over A whereas adding x to A leaves you indifferent between A and B. If indiscriminability entails equidesirability, then this is case where equally desirable options respond differently to equivalent sweetenings.

This will set up the case for my positive theory of desire, and with it the argument for bouletic voluntarism.

8.1 Wanting what's best

According to the “what’s best” theory of desire, to desire p is for p to be the most desirable option among the alternatives in question.⁴⁹ Somewhat more precisely, let Q be a variable that ranges over *alternative sets*—i.e., sets of two or more mutually exclusive propositions. The most straightforward form of the analysis is then as follows:

WHAT’S BEST

‘ S wants p ’ is true in context c iff: for the c -determined Q , either p is the most desirable proposition in Q , or p is the disjunction of all the propositions in Q that are tied for most desirable.

Note: this is an especially simple version of the analysis. For one, it doesn’t build in any of the presuppositional features of ‘want’ that certain “what’s best” theorists have argued for—say the presupposition that both p and its negation are compatible with the subject’s beliefs (Heim 1992). For another, WHAT’S BEST, as stated, fails to validate single-premise closure for ‘want’: it does not follow from the fact that ‘ S wants p ’ and ‘(S knows that) p entails q ’ are true in context that ‘ S wants q ’ is true in context, since q will be the most desirable proposition in Q only if $p = q$. The theory can easily be modified to handle these complexities (see, e.g., von Fintel 1999 for a single-premise closure-validating version of the view). But since such issues are orthogonal to our present concerns—and since the resulting theories are more difficult to think about intuitively—it’s better to stick with the simple formulation.

Because the problem really is quite basic: WHAT’S BEST is a non-starter when it comes to TIE-BREAKING and ASCETICISM. If p and q are equally desirable, then neither is (uniquely) best, and so neither can be such that it but not the other is desired in particular. So discriminating desires in Buridan cases are impossible. And if p and q are *unequally* desirable, then relative to the alternative set $\{p, q\}$ exactly one is best, and so relative to $\{p, q\}$ one cannot fail to have a desire for the better one in particular. So asceticism is impossible. Evidently, then, wanting is not in general a matter of wanting what’s best.

8.2 Wanting what's good enough

So much for the “what’s best” theory. Next we’ll consider the “what’s good enough” theory, which says, roughly, that to desire p is for p to be sufficiently desirable.⁵⁰

Unlike the “what’s best” analysis, which accounts for the facts about desire using purely ordinal comparisons of desirability, the “what’s good enough” analysis is typically developed in a way that requires *cardinal* comparisons of desirability.⁵¹ Now, nothing we’ve said so far about the

⁴⁹ Proponents of this kind of analysis include Lewis (1974), Davis (1984), Stalnaker (1984), Pettit (1991, 2002), Heim (1992), Dreier (1996), von Fintel (1999), Levinson (2003), Condoravdi & Lauer (2016), Jerzak (2019).

⁵⁰ Proponents of this kind of analysis include Bradley (1999), van Rooij (1999), Lassiter (2011), Blumberg & Hawthorne (2021), and Phillips-Brown (2021, 2025).

⁵¹ This isn’t mandatory: van Rooij (1999), for instance, develops a view on which to desire p is for p to be more desirable than the tautology $\top$ (i.e., the status quo). But we’ll soon see reasons to want a more sophisticated implementation.

qualitative notion of desirability guarantees that it has kind of structure that allows for cardinal comparisons. But for the purposes of discussion I’m happy to assume it does. So for the remainder of this section I will follow [Phillips-Brown \(2021\)](#), a proponent of the “what’s good enough” analysis, in understanding desirability in terms of the notion of *expected utility* (in the manner of [Jeffrey 1965](#)). In particular, I will assume that each agent (S) can be associated with a credence function (C_S) and a utility function (U_S). The credence function C_S maps maximally specific possibilities (w) to a real number in the unit interval, representing the subjective likelihood for S that things are as they are in that possibility. (I will assume that C_S obeys the axioms of the probability calculus.) The utility function U_S maps maximally specific possibilities to real numbers, representing the actual subjective value of those possibilities for S . The expected utility function for S (EU_S) is then given as follows:⁵²

$$EU_S(p) = \sum_{w \in W} C_S(w|p) \cdot U_S(w)$$

With desirabilities now modeled as expected utilities, we can regiment the “what’s good enough” theorist’s notion of ‘sufficient desirability’ in terms of *thresholds* of expected utility. This in turn lets us say that for some real number T , you desire p if and only if $EU_S(p) > T$.

What about the value of T —where does that come from? For the most part we’ll be fine without a substantive answer to this question. However, [Phillips-Brown \(2021, 2025\)](#) argues persuasively that it’s at least a context-sensitive matter. This is because going contextualist makes it easy to explain certain patterns in our desire talk, like the fact that a single agent can speak truly in saying both ‘I don’t want any of the dinner options: they’re all mediocre’ and ‘I want the bolognese: it’s the best of the lot’. By allowing for contextual variation in the desirability threshold, we can say that the first utterance is true in a context where the relevant threshold is reasonably high, while the second is true in a context where the threshold is low enough so that the bolognese surpasses it.

So, like Phillips-Brown, I’ll focus on a threshold-contextualist implementation of the “what’s good enough” analysis. Where before we wrote ‘ T ’, we’ll now write ‘ T_c ’ (short for ‘threshold as determined by context c ’). Here, then, is the version of the “what’s good” enough analysis that will be our focus—what Phillips-Brown calls the “Lockean” theory of desire:⁵³

LOCKEANISM

‘ S wants p ’ is true in context c iff: $EU_S(p) > T_c$.

How does the Lockean analysis fare with respect to TIE-BREAKING and ASCETICISM? Well, in the case of TIE-BREAKING it’s no better off than the “what’s best” analysis. The reason why is simple. By hypothesis, the options in Buridan cases are equally desirable (i.e., have equal expected utility) for the subject in question. So either both options are above the threshold or neither is. If they’re both above, then they’re both wanted; if neither is above, then neither is wanted. Either way, it can’t be that exactly one is wanted. But we know that agents can and

⁵² Just as each person is plausibly associated with a family of desirability orderings, I assume that each person is likewise associated with a family of expected utility functions (with differences in the expected utility functions corresponding to differences in one or both of the underlying credence and utility functions).

⁵³ Like the simple version of the “what’s best analysis” sketched above, the version(s) of the Lockean theory I’ll be discussing in this section will abstract away from issues concerning the presuppositions of ‘want’, as well whether ‘want’ obeys single-premise closure. For closure-friendly implementations of Lockeanism, see [Blumberg & Hawthorne \(2021\)](#) and [Phillips-Brown \(2025\)](#).

often do want/prefer exactly one of the options in Buridan cases. So wanting can't just be a matter of surpassing some threshold of expected utility.

Now, it is not at all clear to me how the Lockean account might be amended to handle discriminating desires in Buridan cases. But let's come back to this later; in the meantime it will be instructive to see how the Lockean fairs with respect to asceticism.

At first blush, the Lockean account looks well equipped to handle the possibility of asceticism. In fact, the idea that one's desires can fail to single out the most desirable option is one of the standard motivations for the (contextualist) Lockean theory of desire (Blumberg & Hawthorne 2021, Phillips-Brown 2021, 2025). To account for the natural readings of sentences like 'I know I'd be better off with the free dollar than without it, but it wouldn't be fair to say that I want/prefer the free dollar in particular', all the Lockean needs to say is that such speeches occur in contexts where the expected utility of a free dollar fails to exceed the contextually determined threshold for wanting—i.e., contexts in which $T_c \geq EU_S(\$1)$.

Note, however, that the Lockean needs to do more than account for the possibility of asceticism. They also need to account for the apparent possibility of *interpersonal variation* in asceticism. Again, the central claim of §7 was that people *can* be indifferent to difference; not that they must be. Some people are indifferent to a free dollar, but plenty aren't. So suppose Mr. Apathy is the former kind of person (indifferent to the free dollar) and Ms. Zeal the latter (prefers the free dollar). Both Apathy and Zeal acknowledge that they would be better off having the free dollar (and that no one would be worse off for it), but only Zeal has a preference for it. What is the Lockean's story about the difference between Apathy and Zeal?

Well let's assume for now that both 'Apathy doesn't want the free \$1' and 'Zeal does want the free \$1' are true in a single context c (as it seems they easily could be). It then follows from LOCKEANISM that $EU_Z(\$1) > T_c \geq EU_A(\$1)$. So perhaps the Lockean should say just that: the reason Zeal wants the free dollar but Apathy doesn't is that the two agents assign different expected utilities to the prospect of a free dollar, with only Zeal's assignment clearing the threshold. (Maybe Apathy is considerably wealthier than Zeal, and thus more susceptible to the diminishing marginal utility of money.)

Unfortunately, things aren't so simple. As a warmup, note that Apathy and Zeal both assign greater expected utility to the prospect of a free dollar than to the prospect of nothing. We know Zeal does, because she prefers the dollar to nothing. And it's safe to assume Apathy does, because (i) he can truly say things like 'Of course I'd be better off with the free dollar'; and because (ii) we can imagine the case in such a way that for some number of dollars $1 < n \leq 1,000$, Apathy would prefer $\$1 + \n to $\$0$ but not $\$0 + \n to $\$1$ (recall the discussion at the end of §7). So it looks to follow that:

$$EU_Z(\$1) > T_c \geq EU_A(\$1) > EU_{A/Z}(\$0)$$

However, for reasons exactly analogous to (i) and (ii), we can argue that for any (possibly fractional) positive number n , Zeal assigns greater expected utility to the prospect of a free n cents than to the prospect of nothing—i.e., that for any $n > 0$, $EU_Z(n\text{¢}) > EU_Z(\$0)$. Thus, so long as there is *some* (possibly fractional) positive number of cents n such that the expected utility Zeal assigns n cents is no greater than the expected utility Apathy assigns a dollar—i.e.,

there is an $n > 0$ such that $EU_A(\$1) \geq EU_Z(n\phi)$ —we have that:⁵⁴

$$EU_Z(\$1) > T_c \geq EU_A(\$1) \geq EU_Z(n\phi) > EU_{A/Z}(\$0)$$

And this in turn implies that, for some positive n , Zeal fails to want a free n cents, since for Zeal the expected utility of n cents fails to exceed whatever threshold is responsible for Apathy's apathy to the dollar. Thus, given LOCKEANISM, the following inference is valid: 'It's not fair to say that Apathy wants a free dollar; therefore, there's some number of cents n such that it wouldn't be fair to say that Zeal wants a free n cents'.

But this inference seems terrible. Why should the mere fact that Apathy exhibits some degree of financial asceticism imply that Zeal must also exhibit some degree of financial asceticism (even if it's to a lesser degree)? It's not like Zeal would have to be lying in saying that she would prefer having an extra penny (or a coin-flip for an extra penny) to not having it. So surely Zeal could be such that for any amount of money, no matter how small, she prefers having the free money to not having it. And this means that the explanation of why 'Apathy doesn't want a free dollar' and 'Zeal does want a free dollar' both express truths can't *in general* be that there is some particular threshold T_c such that $EU_Z(\$1) > T_c \geq EU_A(\$1)$.

More generally, it is a consequence of LOCKEANISM that for any two agents S_1 and S_2 and propositions p and q , if $EU_{S_1}(p) = EU_{S_2}(q)$, then ' S_1 wants p if and only if S_2 wants q ' is a context-invariant truth. So even if agents can vary in how much expected utility they get from certain prospects (like that of a free dollar), the Lockean is nonetheless committed to the false claim that if anyone is ascetic to some degree, then *everyone* is ascetic to some degree. A more sophisticated account of asceticism is needed.⁵⁵

On that matter, I think the Lockean's best bet is to treat the possibility of interpersonal variation in asceticism as evidence that the relationship between desire and expected utility is *subject-sensitive* (in addition to context-sensitive). For the Lockean needs there to be true interpretations of speeches like ' S_1 wants p and S_2 doesn't want q ' even when $EU_{S_1}(p) = EU_{S_2}(q)$. And the only way for there to be such interpretations in the Lockean framework is for each subject to be associated with their own expected utility threshold (per context).⁵⁶ This gives us:

SOPHISTICATED LOCKEANISM

' S wants p ' is true in context c iff: $EU_S(p) > T_{S,c}$.

⁵⁴ Given that $EU_A(\$1) > EU_A(\$0)$, the claim that there is *no* (positive) n such that $EU_A(\$1) \geq EU_Z(n\phi) > EU_{A/Z}(\$0)$ is tantamount to the claim that the expected utility Apathy gets from the dollar, though non-zero, is less than the expected utility Zeal would get from a free n cents, no matter how small n is. It is quite hard to see what could motivate this kind of "lexical" difference in the value of money for Apathy and Zeal.

⁵⁵ One option, which I won't explore in detail, is to posit yet further context-sensitivity in the interpretation of 'want'. For instance, one could insist that, holding context fixed, the inference from 'Apathy doesn't want a free dollar' to 'So there is some number of cents n such that Zeal doesn't want a free n cents' is strictly speaking valid. It's just that the natural interpretations of 'Apathy doesn't want a free dollar' and 'There is some number of cents n such that Zeal doesn't want a free n cents' involve different resolutions of the context-sensitivity of 'want', with the natural interpretations of the former involving a more demanding threshold for desire than those associated with the latter.

However, I don't think this proposal has much to recommend it. If there is an interpretation of 'want' on which 'There is some number of cents n such that Zeal doesn't want a free n cents' expresses a truth—which there must be, since there is an interpretation of 'want' on which Apathy would speak truly in saying 'I don't want a free dollar'—then why is it so difficult to access it? And what possibly could explain the difference in the meta-semantic profiles of these sentences? (Is it supposed to have something to do with the meta-semantic profiles of the names 'Apathy' and 'Zeal'?)

⁵⁶ Zyglewicz (forthcoming) defends a subject-sensitive decision-theoretic account of desire, though his is not a version of Lockeanism, and his motivations are pretty different from those I've been pressing here.

(Simple) LOCKEANISM says that S desires p just in case the expected utility S assigns p exceeds the contextually determined threshold of sufficient desirability. SOPHISTICATED LOCKEANISM, by contrast, says that S desires p just in case the expected utility S assigns p exceeds the contextually determined threshold of sufficient desirability *associated with* S . In virtue of this extra degree of freedom, the sophisticated Lockean has a much easier time accounting for interpersonal variation in asceticism.

With respect to Apathy and Zeal, all the sophisticated Lockean needs to say is that Apathy's threshold of sufficient desirability ($T_{A,c}$) is higher than Zeal's ($T_{Z,c}$)—and in particular that:

$$T_{A,c} \geq EU_A(\$1) \geq EU_Z(n\text{¢}) > T_{Z,c}$$

For then it could be that the following three claims are all true (relative to a single context): (i) the expected utility Zeal gets from n cents is no greater than the expected utility Apathy gets from a dollar; (ii) Apathy fails to want the dollar; and yet (iii) Zeal wants the n cents. This is enough to rule out the validity of the inference from 'Apathy doesn't want the free dollar' to 'So there is some positive number of cents n such that Zeal doesn't want a free n cents'—thereby blocking the main challenge to (simple) LOCKEANISM.

As far as accounting for the possibility of interpersonal variation in asceticism goes, SOPHISTICATED LOCKEANISM strikes me as basically adequate. However, it is worth stressing just how significant of a departure it is from the simple version of LOCKEANISM we started with. On the simple, subject-insensitive picture, desire supervenes on desirability. Fix what you mean by 'want'—that is: what the contextually relevant threshold of sufficient desirability is—and you can read off the facts about what S desires from the facts about how desirable S finds the propositions in question. This is not so on the sophisticated picture. On the sophisticated picture, fixing the context doesn't fix the threshold of sufficient desirability, since there is no such thing as *the* threshold of sufficient desirability in context. Each agent has their own threshold (per context). So knowing how desirable S finds p isn't enough to know whether S wants p ; you also have to know what S 's threshold is. And this in turn naturally raises questions about the grounds of a particular agent's threshold. Why, for example, does Apathy have a higher threshold than Zeal?

Of course, explanations have to end somewhere, and this might be one of those places where the Lockean is entitled to appeal to brute facts (at least at the level of folk psychology). But, as a preview of what's to come later, it is also a place where voluntaristic explanations start to look tempting. For one possible explanation of the difference between agents like Apathy and Zeal is in terms of the activity of their wills: the reason Zeal wants the extra dollar while Apathy doesn't is that Zeal made up her mind to have a low threshold while Apathy made up his mind to have a high threshold.

I'm not going to defend this suggestion in any real detail now, since (i) I reject the threshold-theoretic account (even in its sophisticated form) and (ii) I'm not quite ready to make the full-throated case for bouletic voluntarism. So for present purposes the important point is this: *even if* sophisticated Lockeanism were the correct account of the relationship between desire and desirability, we'd still have the relevant kinds of supervenience failures, and with them a natural path to bouletic voluntarism.

All that said, I think the antecedent of this conditional is false: despite its sophistication, SOPHISTICATED LOCKEANISM isn't the correct account of the relationship between desire and desirability. This is for two reasons, the discussion of which will conclude this section.

The first reason is that the Lockean (whether simple or sophisticated) predicts that the relationship between desire and expected utility is *monotonic*: if you desire p when its expected utility is x , then you must desire any proposition whose expected utility is $y \geq x$. Likewise in the other direction: if you fail to desire p when its expected utility is x , then you must fail to fail to desire any proposition whose expected utility is $y \leq x$. But this too seems wrong. To see why, suppose Jones is offered a choice of one of three boxes: the first contains \$100, the second \$200, and the third \$300. And suppose that the relationship between money and expected utility is linear for Jones.⁵⁷ It is a consequence of (sophisticated) Lockeanism that, holding context fixed, the following inference is valid: 'Jones only wants the third box; therefore, if all three boxes contained twice as much money as they actually do, Jones would want the second box too'. Likewise in the other direction: it is a consequence of (sophisticated) Lockeanism that, holding context fixed, the following inference is valid: 'Jones wants the third box; therefore, if all three boxes contained half as much money as they actually do, Jones wouldn't want any of them'. But such inferences are bizarre. They rule out the possibility that Jones simply wants the best box he can get—no matter the absolute value of the contents inside—in which case neither doubling nor halving the value of the boxes will make a difference to what he wants.⁵⁸

The second reason is the one we set aside earlier: any account of desire in terms of thresholds of desirability (whether simple or sophisticated) is powerless to explain discriminating desires in Buridan cases. I've chosen not to belabor this point because I don't think the Lockean's failings here are as instructive as they are for asceticism and monotonicity. But make no mistake: I see tie-breaking as quite a serious problem for the Lockean, and I have no real sense for how the view could be modified to give a plausible account of it.

In light of these issues, I think we have strong reason to reject the idea that desire is a matter of surpassing some threshold of desirability, even when the nature of that relationship is both context- and subject-sensitive. I now turn to defending my preferred account.

9 Cogent desire

9.1 Background assumptions

I'll start with two bits of stage-setting. First, like the "what's best" theorist, I'm going to assume that desire is *alternative-sensitive*—i.e., that desire is a relation between an agent, a proposition, and a set of two or more mutually exclusive propositions (the alternatives under considera-

⁵⁷ This assumption is inessential to the argument. We could instead run a version that quantities abstractly over quantities of money to which Jones is indifferent and quantities of money to which he is not. The concrete numbers just make things easier to think about.

⁵⁸ It would be a mistake to regard this argument as telling meaningfully in favor of WHAT'S BEST, since we can run a parallel version against that analysis too. In particular, the following inference also seems invalid: 'Jones wants the third box in particular; therefore, if each box contained $\frac{1}{100}$ the money it actually does, Jones would still want the third box in particular'. And the reason it seems invalid is that it rules out the possibility that Jones is indifferent to \$3 differences. But given WHAT'S BEST this inference should be perfectly legitimate.

tion).⁵⁹ It's one thing to desire p relative to the alternative set $\{p, \neg p\}$, another to desire p relative to the alternative set $\{p, q, r\}$. By extension, I'm going to assume that one of the dimensions of the context-sensitivity of 'want' (and its cognates) is the alternative set relative to which it is interpreted. These ideas have received considerable support in some of the recent literature on desire—including from Lockeans—so I won't be saying much in their defense here.⁶⁰ By my lights, the best justification for this extra parameter of context-sensitivity is in the explanatory power of the theories that incorporate it.

Second, now that we've dispensed with Lockean thresholds, we no longer need the full structure of expected utility functions to say most all of what we want to say about the relationship between desire and desirability. So for now I'm going to abandon the ideology of expected utility functions (EU_S) and go back to theorizing directly in terms of the qualitative notion of desirability. But it will be useful to formalize the notion a bit.

I'm going to assume that desirability (for S) can be factored into two binary relations over the space of propositions: $>_S$ (pronounced 'is more desirable to S ') and $=_S$ (pronounced 'is equally desirable to S '). ' $p \geq_S q$ ' will abbreviate ' $p >_S q$ or $p =_S q$ '. I will refer to the combination of $>_S$ and $=_S$ as S 's *desirability ordering*.

My assumptions about the logical properties of desirability orderings will be fairly austere—far more austere than they'd need to be if we wanted to represent them with (expected) utility functions. In particular, I'll only be assuming the following:

- $>_S$ is asymmetric (and thus irreflexive). (For any p, q : if $p >_S q$, then $q \not>_S p$.)
- $=_S$ is symmetric, reflexive, and transitive. (For any p : $p =_S p$. For any p, q : $p =_S q$ iff $q =_S p$. And for any p, q, r : if $p =_S q$ and $q =_S r$, then $p =_S r$.)
- $>_S$ and $=_S$ are disjoint. (For any p, q : if $p >_S q$, then $p \neq_S q$.)

So, some salient things I will *not* be assuming about desirability orderings include:

- That $p \geq_S$ is total. (So I'm allowing that for some p, q : neither $p \geq_S q$ nor $q \geq_S p$.)
- That $>_S$ is transitive. (So I'm allowing that for some p, q, r : $p >_S q >_S r \geq_S p$.)

I've chosen to keep the constraints on desirability orderings as trim as they need to be to set up my positive theory of desire and the corresponding argument for bouletic voluntarism. By not requiring $\geq_S$ to be total, we leave room for the possibility of desirability-incomparability (of the kind briefly discussed in §3.2 above).⁶¹ And by not requiring $>_S$ to be transitive, we leave

⁵⁹ Note that I am not assuming that the alternative set will always be a *partition*. The alternative sets to which desire is sensitive can fail to exhaust logical space.

⁶⁰ For some recent work in support of the alternative-sensitivity of desire, see, e.g., [Blumberg & Hawthorne \(2021, 2022\)](#), [Blumberg \(2022a\)](#), [Grano & Phillips-Brown \(2022\)](#), [Phillips-Brown \(2025, forthcoming\)](#), [Zyglewicz \(forthcoming\)](#). Of course, by introducing alternative-sensitivity into one's theory of desire, one opens oneself up to a number of difficult questions about the interpretation of 'want' (and its cognates). Chief among them is the question of what, meta-semantically speaking, determines which alternative set is operative in a given context. But rather than try to resolve these questions, I'll simply follow the precedent set by the literature and leave them as a task for future research.

⁶¹ See also [Levi \(1986\)](#), [Gert \(2004\)](#), and [Rabinowicz \(2008\)](#).

room for Rachels-Temkin-style responses to “spectrum” arguments.⁶²

That said, unless stated otherwise, every example I discuss will be one in which it is assumed that the agent in question has a desirability ordering that satisfies these last two assumptions: that is, for the relevant agent S , $\geq_S$ is total and $>_S$ is transitive (at least for the propositions under consideration). This assumption will allow us to express the deliverances of desirability orderings more compactly, as in:

$$p =_S q >_S r >_S s$$

Rather than the cumbersome:

$$p =_S q, p >_S r, p >_S s, q >_S r, q >_S s, r >_S s$$

And it will also allow us to reason intuitively about how the agent’s desires are liable to be affected by changes to the desirability of the options under consideration. But again: the paper’s “official” conception of desirability is the logically austere one.

With that, we are now in a position to state the core principle underlying my theory of desire. I’ll introduce it with an example.

9.2 Cogency

Suppose S is thinking about which restaurant to go to tonight. They’ve narrowed it down to a choice between A, B, C, and D. The only relevant consideration is S ’s assessment of the desirability of the restaurants, which they give voice to as such: ‘I love all four of these restaurants; but if I had to rank them I’d say A and B are tied for best, with C a close third and D a close fourth’. Let us suppose, then, that S ’s desirability ordering is as follows:

$$A =_S B >_S C >_S D$$

Suppose we ask S which restaurant they want to go to tonight. How might they respond? Here, to my ears, are some of the answers we could expect them to give, holding fixed our knowledge of their ranking of the options:

- A. ✓
- B. ✓
- A or B. ✓
- Not D. ✓
- Either A, B, C, or D / Any of them will do. ✓

And here are some of the answers that would be strange for S to give:

⁶² See, e.g., [Temkin \(1987, 1996\)](#), [Rachels \(1998, 2001\)](#). For critical discussion, see, e.g., [Handfield \(2014\)](#), [Nebel \(2018\)](#). There is also a sizable literature on the question of whether “preference” is transitive, and the consequences of it failing to be—see, e.g., [Davidson *et al.* \(1955\)](#), [Quinn \(1990\)](#), [Hansson \(1993\)](#), [Gendin \(1996\)](#), [Gustafsson \(2016\)](#), and [Baumann \(2022\)](#). Note, however, that this literature often understands ‘preference’ in a quasi-theoretical way, operationalized in terms of an agent’s choice dispositions. As such, the exact relationship between the non-transitivity of (this kind of) preference and the non-transitivity of $>_S$ seems to me to be largely an open question. We’ll return to some of these issues in the final few sections of the paper.

- C. ✗
- Any of them besides B. ✗
- A or D. ✗

What explains this pattern of judgments? Well, intuitively, the strange-seeming answers seem strange because they all involve *S* wanting options they know to be less desirable than at least some of the available alternatives. In saying ‘C’, *S* is representing themselves as wanting C in particular. But why would they want C in particular, given that they know that A and B are better? Likewise, in saying ‘A or D’, *S* is representing themselves as being unpicky enough to be happy with either A or D, but picky enough to be unhappy with either B or C. But if they’re going to be picky enough to rule out B and C, why wouldn’t they be picky enough to rule out D? Or, in the other direction: if they’re going to be unpicky enough to include D, why wouldn’t they be unpicky enough to include B and C? More generally, if *S* is willing to settle for two of the options, it seems like it should be the two most desirable ones: A and B. If they’re going to report themselves as desiring in a way that contravenes this idea, it strongly suggests they’re either lying about what they want, or that their desirability ordering is other than what was suggested by their initial speech about the relative quality of the restaurants.

The reasonable seeming answers, by contrast, all play nicely with the sorts of considerations discussed in §§6-7. It’s reasonable to express a desire for A in particular or for B in particular, because it’s reasonable to have discriminating desires between equally desirable options, and A and B are tied for most desirable. But it’s also reasonable to express a desire not for A in particular or for B in particular, but instead for their disjunction, since it’s also reasonable not to have discriminating desires between equally desirable options. And it’s reasonable to have a desire for anything but D, or to have no preference at all between the four options, since it’s reasonable to be an ascetic.

So what’s the general principle here? If the arguments of §8.2 are correct, then it’s not the Lockean’s: the relationship between desire and desirability cannot simply be a matter of surpassing some threshold (even if that threshold both context- and subject-sensitive). Instead, the core principle seems to be that one’s desires are *cogent* with respect to the facts about desirability, in a sense analogous to which one’s guesses are cogent with respect to the facts about subjective likelihood (Holguín 2022, Dorst & Mandelkern 2023). Indeed, given the similarities in the patterns here and the patterns about guessing discussed in §4, this connection seems quite natural.

Put roughly, the idea behind cogency is this. Given a set of alternatives *Q*, you’re allowed to be as picky or unpicky as you please with respect to which of the options in *Q* are compatible with what you desire (relative to *Q*). But your pickiness is constrained by the desirability of the options under consideration. If you are so picky as to desire a single alternative, it better be best or at least no worse than any other; and if you are unpicky and so indifferent between some number of alternatives, the each they need to be no less desirable than the alternatives that are *not* compatible with what you desire.

Somewhat more precisely, say that a proposition *p* is *defined* on a set of alternatives *Q* just in case *p* is equivalent to some union of cells of *Q* (i.e., *p* is either one of the propositions in *Q*

or is a disjunction of two or more propositions in Q).⁶³ We can then say:⁶⁴

COGENCY

p is cogent for S relative to Q iff:

1. p is defined on Q .
2. For every $q \in Q$: if q entails p , then: for every $r \in Q$ such that $r >_S q$: r entails p .

Suppose that S 's desires relative to $\{A, B, C, D\}$ are cogent in the sense just defined. It would then follow that: relative to $\{A, B, C, D\}$, and holding fixed S 's desirability ordering, S desires one of the following five propositions:

- A
- B
- $A \vee B$
- $A \vee B \vee C$
- $A \vee B \vee C \vee D$

I submit that this exactly matches the intuitive judgments about what S can be reasonably expected to desire in the dinner case. S might be picky and desire A (B), because there is no option incompatible with that desire that is better than A (B). Or S might merely desire $A \vee B$, being picky enough to narrow things down to the two (tied-for)-best options, but not picky enough to form a Buridan preference. Or S might exhibit some degree of asceticism—from the partially ascetic desire for $A \vee B \vee C$ to the fully ascetic desire for $A \vee B \vee C \vee D$. All cogency guarantees is that if S 's desire is compatible with exactly n of the alternatives, then they are the n -best alternatives (give or take ties), as judged by S 's desirability ordering.⁶⁵

From here on out I will be assuming that the only necessary connection between desire and desirability is given by cogency. I know of no other principle that allows desire to be constrained by desirability without ruling out tie-breaking and asceticism. And I submit that all the examples discussed in the paper so far accord nicely with the hypothesis that to desire is to desire cogently. The task for the rest of the section is to present a theory of desire that captures these ideas.

9.3 Desire in terms of cogency

Cogency is a permissive condition. Even supposing there is only one answer to the question 'How should S 's desirability ordering rank each of the propositions in Q ?', there will always be more than one answer to the question 'What's cogent for S relative to Q ?'. In the limiting case where

⁶³ Note that as 'defined' is defined, a *conjunction* of two or more propositions in Q is not defined on Q .

⁶⁴ The cogency(/"filtering") principle for guessing defended by Holguín (2022) and Dorst & Mandelkern (2023) is essentially identical to the principle I'm calling COGENCY: just replace 'desire' with 'guess' and ' $>_S$ ' with 'is assigned higher credence by S ' and you get the guessing principle. Or almost, anyway: both Holguín and Dorst and Mandelkern state a version of the principle that requires one's guess to include every alternative r such that: S assigns *no less* credence to r than to any alternative q included in one's guess. This is tantamount to not allowing one's guesses to break credal ties. Holguín doesn't offer any justification for this constraint, while Dorst and Mandelkern explicitly acknowledge that it might be appropriate to relax the condition to allow for tie-breaking (see their footnote 8). For the reasons given in §4, it seems clear that a good theory of rational guessing will allow for arbitrary tie-breaking.

⁶⁵ Note that if $>_S$ is non-transitive and/or $\geq_S$ is incomplete with respect to the options in Q , this gloss will be a bit misleading. In such cases it would be better to think about S 's Q -relative desires as being determined not by which options are the n -best (for the relevant $1 \leq n \leq |Q|$), but which options are the $(|Q| - n)$ -least bad.

$Q = \{p, q\}$ and $p >_S q$, there are at least two cogent _{S, Q} propositions: p and $p \vee q$.⁶⁶ Adding more propositions to Q (and/or introducing desirability ties and/or incomparabilities) only increases the space of cogent possibilities. The restaurant case has five such possibilities, for example. In light of this, we know at a minimum that a proposition's being cogent cannot *suffice* for its being desired.

However, there are also good reasons to think that cogency cannot be a *necessary* condition on desire either. Consider the restaurant case, but with another character, S^* , whose desirability ordering is as follows:

$$C >_{S^*} B >_{S^*} D >_{S^*} A$$

Suppose S says they want A and that S^* says they want C . In such circumstances, a reasonable answer to the question 'How many people are such that they want A or C ?' is 'Two'. Likewise, we can reasonably assert things like ' S and S^* are alike in that they both want A or C '. It would thus seem reasonable to conclude that both S and S^* want _{$\{A, B, C, D\}$} $A \vee C$.⁶⁷ And this is despite the fact that, relative to $\{A, B, C, D\}$, $A \vee C$ is neither cogent for S nor cogent for S^* . This suggests that we can desire non-cogently.

More generally, the idea that cogency is necessary for desire plays badly with the plausible (albeit controversial) idea that desire obeys single-premise closure: i.e., that if (i) S desires _{Q} p and (ii) p entails q , then (c) S desires _{Q} q .⁶⁸ This is for the basic reason that cogency _{S, Q} doesn't obey single-premise closure: it does not follow from the fact that A is cogent _{$\{A, B, C, D\}$} and that A entails $A \vee C$ that $A \vee C$ is cogent _{$\{A, B, C, D\}$} (it's not). So anyone who thinks desire obeys single-premise closure has to give up the idea that our desires are invariably cogent.

Now, the question of whether desire obeys single-premise closure is hotly contested.⁶⁹ It thus needn't be such a disappointment to have a theory of desire that fails to validate it. However, I myself am sympathetic to the validity of single-premise closure for desire, and am persuaded by the suggestion, often found in the literature, that the putative counterexamples to its validity involve non-uniform resolutions of the context-sensitivity of 'want' (thanks to shifts in the alternative set and/or underlying desirability ordering). But since questions of the logic of desire are largely orthogonal to our main line of inquiry—and since it makes my job strictly harder if I have to validate single-premise closure than if I don't—I'll simply be taking its validity for granted in what follows.

So: how can we build a theory of desire out of the idea that desire is governed by cogency, given that *being cogent* is neither necessary nor sufficient for *being desired*? The answer is by introducing a new theoretical primitive. In particular, I'm going to assume that for any agent S and alternative set Q , there is always a *strongest* cogent proposition S desires relative to Q .⁷⁰

⁶⁶ For notational convenience, I'll write 'desire _{Q} ' as shorthand for 'desire relative to Q ', 'cogent _{S, Q} ' for 'cogent for S relative to Q ', etc.

⁶⁷ Such a conclusion is not mandatory, however. It could be that the true interpretation of this speech is more perspicuously expressed by something like ' S and S^* are alike in that they both are such that either they want A or want C '. (*Mutatis mutandis* for the readings of the question 'How many people are such that they want A or C ?' on which 'Two' is a reasonable answer.) See footnote 69 below for citations of some of the relevant literature here.

⁶⁸ I'm happy to leave open what the relevant notion of *entailment* amounts to (logical? metaphysical? known/believed logical/metaphysical?). For our purposes any reasonable notion will work well enough.

⁶⁹ See, e.g., Heim (1992), von Fintel (1999), Crnić (2011), Braun (2015), Blumberg & Hawthorne (2021, 2022), Blumberg (2022a,b), Booth (2023), Phillips-Brown (2025), and Zyglewicz (forthcoming) for relevant discussion.

⁷⁰ Strictly speaking, it would be better to weaken this claim so that it applies only to agents who have *some bouletic*

Somewhat more precisely: for any S and Q , there is a p such that: (i) S wants _{Q} p ; (ii) p is cogent _{S,Q} ; and (iii) for any proposition q , if S wants _{Q} q , then p entails q . Whichever propositions satisfies (i)-(iii) is the content of (what I will call) S 's *ur-desire* _{Q} .

With the notion of *ur-desire* in hand, we can give a simple, cogency-theoretic analysis of desire as follows:

COGENT DESIRE

S wants _{Q} p iff: p is entailed by S 's *ur-desire* _{Q} .

COGENT DESIRE does not make cogency sufficient for desire: of the various cogent _{S,Q} propositions, at most one is the content of S 's *ur-desire* _{Q} , and only the propositions that are entailed by the content of that *ur-desire* _{Q} are desired _{Q} . Nor does it make it necessary for desire: propositions entailed by one's *ur-desire* _{Q} aren't guaranteed to be cogent _{S,Q} , but they are guaranteed to be desired _{Q} . Nonetheless, COGENT DESIRE does justice to the idea—brought out by the arguments of §§6-8—that the essential nature of the relationship between desire and desirability is given by cogency. Non-cogent desire is possible, but it's always *derivative* desire; relative to any particular alternative set, one's (logically) strongest desire is invariably cogent.

COGENT DESIRE is more or less the paper's official theory of desire. As far as it is concerned, any cogent _{S,Q} proposition could constitute S 's *ur-desire* _{Q} . This guarantees that S can have discriminating desires in Buridan cases, and can exhibit any degree of asceticism. It also guarantees that S desires _{Q} the consequences of what S *ur-desires* _{Q} , ensuring that S 's desires _{Q} obey single-premise closure.

Beyond the constraints imposed by cogency and single-premise closure, the theory's predictions are sparse. Knowing that $A =_S B >_S C >_S D$ is enough to know that S has one of five possible *ur-desires* _{$\{A,B,C,D\}$} ; but that's about it. In the other direction, knowing that S desires _{$\{A,B,C,D\}$} A is enough to know that S finds A no less desirable than any of B , C , and D ; but that's about it. However, it's also not clear that we should *want* our theory of desire to have more predictive power than this. After all, the main lesson of §§6-8 is that there is considerable intrapersonal leeway in which states of desire one *could* adopt, and considerable interpersonal leeway in which states of desire agents *do* adopt, even in cases where the relative desirability of the options under consideration is held fixed.

That said, the lack of predictive power would perhaps be more palatable if we had a more interesting story to tell about the theory's central component: *ur-desire*. For as it currently stands, the theory is plainly circular: desire is analyzed in terms of *ur-desire*, which itself is partially analyzed in terms of desire, and is otherwise simply being treated as a theoretical primitive. The goal of §10 will be to ameliorate this situation, by taking a stand on the metaphysical basis of *ur-desire*. This is where bouletic voluntarism will finally enter the picture.

But first, it will be instructive to examine in greater detail some of COGENT DESIRE's substantive predictions. For even continuing to treat *ur-desire* as an unanalyzed primitive, there is much to scrutinize in the claim that desire is a consequence of (cogent) *ur-desire*. (Note: I encourage impatient and/or exhausted readers to skip to §10 and return here at their leisure. None of the

attitude or other relative to Q , since plausibly desiring p relative to Q requires that (at some level of abstraction) S comprehends p and the alternatives in Q . But I'll suppress this complexity in what follows.

objections discussed below will have much impact on the shape of the discussion in the final few sections of the paper.)

9.4 Putative failures of cogency

Here I will discuss three problem cases for COGENT DESIRE, ordered from least to most concerning. They are: (i) cases in which we seem to want *none* of the options under consideration; (ii) cases in which we seem to want *multiple* (incompatible) options under consideration; and (iii) cases in which we seem to prefer less desirable options to more desirable options.

9.4.1 Wanting none

It is a consequence of COGENT DESIRE that, relative to any alternative set Q , there is at least one option in Q such that: one's desires $_Q$ are guaranteed to be satisfied so long as one gets that option.⁷¹ Equivalently: there will always be at least one option in Q such that that option is compatible with what one ur-desires $_Q$. For example, if $Q = \{p, q\}$, with $p >_S q$, then one's ur-desire $_Q$ is guaranteed to be either p or $p \vee q$ —in which case getting p is a way of satisfying one's ur-desire $_Q$, no matter which of the two possible ur-desires $_Q$ one has.

But we sometimes talk as if *none* of the options under consideration is wanted. For example, if I ask you, 'Which do you want: to spend a month in prison or be tortured for a day?'—'Neither' is a perfectly reasonable response.⁷² Likewise, we can imagine the restaurant case above as one in which S has the same desirability ranking, but where now all the options are considered quite bad in "absolute" terms. In such circumstances, it would be perfectly reasonable for S to say 'I want none of A, B, C, or D; they're all terrible'.

However, I believe this challenge is superficial. The felicity of the "none of the above"-type desire-reports is just further evidence of the context-sensitivity 'want'. Indeed, to account for these uses, all we need to do is allow for resolutions of the context-sensitivity of 'want' on which "none of the above" is included among the contextually salient alternatives. Say that $\mathbf{X}_Q$ corresponds to the negation of the disjunction of all the propositions in the alternative set Q . Then we can say that in contexts where 'None of them' is a natural response to the question 'Which restaurant do you want to go to?', the relevant alternative set is $\{A, B, C, D, \mathbf{X}_{\{A, B, C, D\}}\}$.⁷³ And relative to *that* alternative set, an ur-desire for $\mathbf{X}_{\{A, B, C, D\}}$ (i.e., none of A, B, C, or D) is cogent so long as the desirability of $\mathbf{X}_{\{A, B, C, D\}}$ is no less than the desirability of any of the restaurant prospects (as we can easily imagine it being).

9.4.2 Wanting multiple

According to COGENT DESIRE, for any alternative set Q , there is a unique, strongest proposition S desires relative to Q , which is either one of the alternatives in Q or some disjunction of those alternatives. This proposition is the content of S 's ur-desire $_Q$, and all of S 's desires $_Q$ stem from it. An immediate consequence of this fact is that S 's desires $_Q$ are guaranteed to be *consistent*.

⁷¹ If Q has a unique most desirable option, then that option is the witness to the claim; if there's a tie for most desirable, then at least one of those options will be a witness to the claim.

⁷² Phillips-Brown (2021, 2025) uses these sorts of judgments to argue against WHAT'S BEST and in favor of LOCKEANISM.

⁷³ It could also be $\{\mathbf{X}_{\{A, B, C, D\}}, \neg \mathbf{X}_{\{A, B, C, D\}}\}$.

After all, if p and q are inconsistent, then it cannot be that both p and q are entailed by what S *ur-desires* _{Q} . Instead, S can at most want _{Q} *either* p or q , as when both p and q entail (rather than are entailed by) S 's *ur-desire* _{Q} . In the restaurant case, for example, S cannot both want _{$\{A,B,C,D\}$} A and want _{$\{A,B,C,D\}$} B , since no possible *ur-desire* _{$\{A,B,C,D\}$} entails both A and B . At most S can want _{$\{A,B,C,D\}$} either A or B (as when S 's *ur-desire* _{$\{A,B,C,D\}$} is one of: A , B , or $A \vee B$).

What this means, effectively, is that if COGENT DESIRE is right, then relative to any single alternative set Q , it's impossible to have conflicting desires _{Q} . At most it can be that multiple conflicting propositions are *compatible* with what one desires _{Q} (i.e., are ways of satisfying what one desires _{Q}). And on the face of it this prediction might seem implausible. Isn't it just a banal platitude that our desires can conflict?⁷⁴

Well, perhaps. But care is needed in saying what exactly this platitude amounts to. 'Want' is highly context-sensitive. So the mere fact that we can speak truly in saying 'I want p ' and 'I want q ' even when (we know that) p and q are incompatible is not particularly strong evidence that we can simultaneously want p and want q . For an alternative explanation is that the true interpretations of the conflicting desire reports involve shifts in the contextually relevant alternative set and/or desirability ordering (Levinson 2003, Crnić 2011, Phillips-Brown forthcoming). Indeed, this seems to me the natural diagnosis of what's going on with the "conflict" brought out by (i) 'I want to pay my taxes because I don't want to go to jail' and (ii) 'I don't want to pay my taxes because I'd rather spend the money on frivolities'. In saying both of these things, I am not expressing a single conflicted state of mind, but two distinct states of mind.

However, there are other putative cases of conflicting desire that seem less amenable to this style of contextualist treatment. Indeed, these sorts of cases have been taken to constitute some of the main considerations in favor of the Lockean analysis of desire (Blumberg & Hawthorne 2021, Phillips-Brown 2021, 2025).⁷⁵ Consider, for example, the Buridan case of §6. You're faced with a choice between a \$20 bill in a box labeled 'Left' and a \$20 bill in a box labeled 'Right'. You're asked which of them you want. To my ears, a response along the lines of 'I want both of them' seems perfectly reasonable. Likewise for 'I want Left and I want Right, though I realize I can only have one'—that too seems just fine. But unlike in the taxation case, here it seems quite implausible that the true interpretations of these reports involve non-uniform resolutions of the context-sensitivity of 'want'. For reasons stressed in §6, no reasonable desirability ordering will distinguish the two options. And though we could *try* to jerry-rig some alternative sets to make the judgments come out COGENT DESIRE-compatible—say by taking 'I want both' to be uniformly true relative to $\{\text{Left} \wedge \text{Right}, \text{Neither}\}$, or to involve some kind of non-uniform truth relative to $\{\text{Left}, \neg\text{Left}\}$ and $\{\text{Right}, \neg\text{Right}\}$ —this strikes me as a measure of last resort.

So my preferred response is to distinguish between two kinds of desire—a distinction which is evidence of a deeper kind of context-sensitivity in 'want' than we've seen so far. In particular, I am going to posit a distinction between *ordinary* desire—which is the kind I take myself to have been theorizing about throughout the course of the paper—and what I will call *inclusive* desire (and will write as ' $\cap$ -desire')—which is the kind I claim is often expressed when we report

⁷⁴ For further motivation see, e.g., Greenspan (1980), Pettit (1991), Marino (2009), Baker (2010), Phillips-Brown (forthcoming), as well as the citations in the main text below.

⁷⁵ Given LOCKEANISM, if both p and q are above the (contextually relevant) threshold of sufficient desirability, then both are wanted. Considerations of co-satisfiability are irrelevant.

“conflicting” desires.

I lack the space to make a comprehensive case for an ambiguity treatment. But as a suggestive point, suppose we add a third “Middle” box to our Buridan case, which contains some pocket lint (Left and Right still each contain a crisp twenty dollar bill). Suppose you’re asked ‘You’re about to get one of the three boxes, which of them do you want?’. To my ears the following response is perfectly reasonable:

‘Both Left and Right, just not Middle.’

But so too is:

‘Either Left or Right, just not Middle.’

Does one express a different mental state depending on which of these two answers one gives? I’m inclined to think not. I’m inclined to think that—even though the first answer involves the words ‘both’ and ‘and’ while the second involves the words ‘either’ and ‘or’—one who utters either answer expresses the same mental state (at least in the setting of this particular case). Moreover, supposing you do answer in the first way (‘Both Left and Right, just not Middle’) and get Right, does it follow that you didn’t get everything you wanted? Maybe in one sense yes—a sense we can give voice to with a speech like:

‘You said you wanted both Left and Right, and you only got Right, so you didn’t get everything you wanted.’

But there seems also to be a clear sense in which the inference is unwarranted—a sense we can give voice to with a speech like:

‘You said you just didn’t want Middle, and we gave you something other than Middle, so you got everything you wanted.’

To my ears, neither speech mischaracterizes the underlying bouletic facts.

I find it difficult to make sense of all this on the assumption that ‘want’ is interpreted uniformly across these different speeches. So I’ll simply be taking it for granted that the kind of desire we express when we say things like ‘I want both Left and Right, not Middle’ is of a different kind than the one we express when we say things like ‘I want either Left or Right, not Middle’ (again, in the setting of this particular case). The former is the as-of-yet unanalyzed $\cap$ -desire; the latter is (ordinary) desire. The question is the relationship between these two kinds of desire.

On my preferred analysis, $\cap$ -desire is essentially the mirror-image of (ordinary) desire. Getting what you want (in the ordinary sense) can be as hard as getting the most desirable option (as when you’re maximally picky), and can be as easy as getting any option (as when you’re maximally ascetic). By contrast, getting what you $\cap$ -want can be as easy as getting the most desirable option (as when you’re maximally $\cap$ -picky), and can be as hard as getting all the options (as when you’re maximally $\cap$ -ascetic). And just as we’ve analyzed desire in terms of ur-desire and cogency, so too can we analyze $\cap$ -desire in terms of mirror-image notions of ur- $\cap$ -desire and $\cap$ -cogency.

More precisely:

- I'll say that a non-empty $Q^* \subseteq Q$ is $\cap$ -**cogent**_{S,Q} iff: for each p in Q^* and each q that's in Q but not in Q^* : $p \not\prec_S q$. (By contrast: cogency_{S,Q} is a property of *propositions* defined on Q , to the effect that every Q -alternative that's entailed by the proposition in question is no less desirable than any Q -alternative that is not entailed by it.)
- I assume that for any S and Q , S has an **ur- $\cap$ -desire**_Q **state**, which is a particular non-empty subset of Q . These are all the propositions that S ur- $\cap$ -desires_Q. (By contrast, S 's ur-desire_Q is a particular *proposition* (defined on Q): namely, the disjunction of all the elements of some particular non-empty subset of Q ; namely all the propositions compatible with what S ur-desires_Q.)
- I assume that S 's ur- $\cap$ -desire state is $\cap$ -cogent_{S,Q}.

The analysis of inclusive desire is then as follows:

COGENT $\cap$ -DESIRE

S $\cap$ -desires_Q p iff: there is a q in S 's ur- $\cap$ -desire_Q state such that q entails p .

So, for example, if S 's ur- $\cap$ -desire_{A,B,C,D} state is $\{A, B\}$, then S $\cap$ -desires_{A,B,C,D} A , B , and any proposition that is entailed either by A or by B .

In terms of its predictions, this analysis is in certain respects quite similar to Phillips-Brown's (2025) Lockean theory of desire. In fact, the only important difference is that on Phillips-Brown's analysis, every proposition in Q whose expected utility for S exceeds some contextually salient threshold is included in S 's ur- $\cap$ -desire state. My analysis imposes no such requirement. And for reasons exactly analogous to those given in §8.2, I think we do better *not* imposing this requirement, since it effectively rules out discriminating $\cap$ -desires in Buridan cases, as well as interpersonal variation in levels of $\cap$ -asceticism. And both phenomena seem as possible for conflicting desire as they do for ordinary desire.⁷⁶

Moreover, to make good on the suggestion that (ordinary) desire and $\cap$ -desire are mirror images of one another, all we need is the following bridge principle:

DESIRE / $\cap$ -DESIRE DUALITY

For any $p \in Q$: S ur- $\cap$ -desires_Q p iff p entails S 's ur-desire_Q.

In words: for a proposition in Q to be ur- $\cap$ -desired_Q is for that proposition to be *compatible* with what you ur-desire_Q. So, for example, someone who ur- $\cap$ -desires_{Left, Middle, Right} both Left and Right is someone who ur-desires_{Left, Middle, Right} that they get either Left or Right. Understood as such, desire and $\cap$ -desire are two sides of the same coin.

In light of this, I believe that cases of conflicting desire are not in any real tension with COGENT DESIRE. In some cases, we can account for the judgments using run-of-the-mill contextualist machinery (i.e., shifts in the desirability ordering and/or alternative set). In other cases, we can account for the judgments by taking 'want' to express $\cap$ -desire. From there we can smoothly

⁷⁶ The account is is even more similar to Blumberg & Hawthorne's (2021, §4.1) analysis, since on their view (what plays the role of) S 's ur- $\cap$ -desire state is determined by an unanalyzed "good-selecting function" (which returns a non-empty subset of Q). However, Blumberg and Hawthorne don't appeal to considerations of Buridan cases or interpersonal variations in asceticism to justify their unanalyzed good-selecting function, and in fact express some sympathy with interpreting this function in terms of expected utility thresholds, as Phillips-Brown does.

read off the facts about which of the conflicting propositions the agent $ur\text{-}\neg$ -desires (and thus $\neg$ -desires) from the facts about which propositions are compatible with what the agent ur -desires, or vice versa. Either way, the fundamental principle governing the relationship between desire and desirability is given by cogency.

9.4.3 Wanting worse

COGENT DESIRE says it is possible to fail to want what's most desirable. This is because COGENT DESIRE licenses asceticism, and asceticism is a way of failing to want what's most desirable. But COGENT DESIRE does not say it is possible to want what's *less desirable*. This is because it follows from COGENT DESIRE that for any Q and any $p, q \in Q$, if p but not q is compatible with what S ur -desires $_Q$, then $p \not\prec_S q$. Thus, COGENT DESIRE sharply distinguishes asceticism—failing to desire the more desirable—from *masochism*—actively desiring the less desirable.⁷⁷

Is it appropriate to draw such a sharp line between asceticism and masochism? Here matters turn on subtle issues concerning the theoretical notion of 'desirability', which is still basically just a theoretical primitive. Unsurprisingly, this makes it a bit unclear how we're supposed to go about settling a question like this. Still, I think there are a number of things to be said in favor of drawing this distinction.

Start with the analogy to guessing. Suppose you're sure that a coin is more likely to land heads than tails. It would not be particularly puzzling for you to report yourself as failing to think that the coin will land heads. This is because there is nothing puzzling about being the kind of person who is cautious about venturing a guess, even when you're sure that a certain answer is most likely to be correct. What would be puzzling, however, is if you reported yourself as actively thinking the coin will land *tails*. If you're genuinely sure that the most likely outcome is heads, then in what sense could you really think the coin will land tails? One's comparative confidence judgments may not *determine* one's guesses; but they seem to impose enough discipline to rule out this kind of incongruity. The claim is that the relationship between desire and desirability works in essentially the same way.

Can this claim be given direct intuitive support? On the one hand, there does seem to be an intuitive contrast between a speech like:

'I know A and B are in every relevant respect slightly more appealing than C; but the difference is small enough that I don't care which of the three I go with.'

and a speech like:

'I know A and B are in every relevant respect slightly more appealing than C; but I want to go with C rather than with A or B.'

The latter is puzzling in the way the former is not. The hypothesis that desire must be cogent neatly explains this contrast—at least assuming that an agent's judgments of "appeal" are a reasonable guide to workings of their desirability ordering.

On the other hand though, it's not so difficult to find cases that might reasonably be thought to demonstrate the possibility of masochistic desire. Consider a putative case of *bald-faced*

⁷⁷ To be clear: I am using 'masochism' in a technical sense. It is entirely compatible with everything I say here that masochists in the ordinary sense desire/prefer in ways that accord with their desirability orderings.

masochism. Suppose *S* strongly desires to smoke, all the while recognizing that it is very much in their interest not to. It would not be obviously misguided to describe a case like this as one in which, relative to {smoke, ¬smoke}, *S* strongly prefers smoking, despite finding not smoking more desirable than smoking. *Mutatis mutandis* for others desires that might naturally be described as ill-conceived, compulsive, self-destructive, or what have you: squint and you can see the mark of such cases as the agent desiring the less desirable.

However, I think we should resist this characterization of the cases, and with it the very possibility of bald-faced masochism. Again, ‘desirability’ is a quasi-theoretical notion, understood as the kind of value relevant to desire, *whatever that value may be*. This kind of value will often have substantive connections to one’s well-being, long term goals, and so on; but it is not meant to be inextricably tied to such things. And in cases where an agent is willing to go to lengths to satisfy a certain desire, we have good reason to think that the object of that desire has a kind of value for them deserving of the label ‘desirable’—albeit a kind of value that can come apart from their perhaps more reflective conceptions of the good.⁷⁸ In other words, then, I am happy to simply assume that it’s part of the theoretical role of *desirability* that bald-faced masochism of this sort is impossible. The person who (strongly) prefers to smoke, even while recognizing how much better off they’d be if they didn’t, is a person for whom smoking is more desirable than not, their better judgment notwithstanding.

10 Bouletic voluntarism

Desire is determined by ur-desire, which is in turn constrained but not determined by desirability. But what, if not desirability, settles the facts about ur-desire?

At long last we turn to the matter of the will. My thesis is that we often quite literally choose what we ur-desire, and thus, by extension, what we desire. Bouletic voluntarism is the natural consequence of desire’s failure to supervene on the desirable, and provides the grounds of ur-desire in a non-trivial range of cases.

This goal of this section substantiate and refine these claims. As we’ll see, most of the argumentative work has already occurred; it doesn’t take much effort to repurpose the ideas discussed in §§6-9 into a direct argument for bouletic voluntarism. But the truth of bouletic voluntarism leaves open a number of important residual questions. Section §11 will address three of them. First, if we’re warmed up to thinking our desires are *sometimes* formed at will, why not think they *always* are? Second, supposing that our desires are *not* always formed at will, then what, if anything, can be said about the difference between the cases where our desires are formed at will from the cases where they are not? And third, what are the grounds of non-voluntary (ur-)desire?

But this section will stick to making the case for bouletic voluntarism. I’ll begin indirectly, by considering and arguing against some of the natural objections to the thesis.

⁷⁸ We could, of course, take this to be a *rational* failing on *S*’s part. I’ll have more to say about the rational norms on desirability in §13.

10.1 Against the case against bouletic voluntarism

Why think desires cannot be formed at will?

Here is one kind of thought, implicit in much of the existing literature on desire, and a direct consequence of the “what’s best” and “what’s good enough” theories discussed in §8. Desire aims at the desirable: our desires are formed in response to the desirability facts. We simply *notice* (in a broad sense) that certain things have certain desirability-promoting properties—they appear good, or conducive to pleasure, or to be the sorts of things we have reason to pursue—and, in virtue of noticing these things, come to desire what we do. As a result, there is simply no room for the will to enter the picture when it comes to desire formation. Indeed, if the will *could* enter the picture—if we could simply choose whether to have a certain desire—then desire couldn’t be the kind of thing that aims at the desirable.

However, we know from §§6-7 that this argument is unsound. Our desires are not merely byproducts of the desirability facts. If they were, both discriminating desires in Buridan cases and interpersonal variation in levels of asceticism would be impossible. Yet we know that these things are not just possible, but actual. Of course, this is not to deny that our desires are *constrained* by our judgments of desirability. But it is perfectly compatible with bouletic voluntarism that our patterns of desire formation are, in a law-like way, responsive to the underlying desirability facts. It’s just that the true law—cogency—leaves room for choice.

Guessing once again provides a useful analogy. There is an influential argument, due to Bernard Williams (1973, ch. 9), against the possibility of believing at will. In crude outline, the argument is this. Belief “aims at truth”: it is constitutive of belief that when one believes something, one takes it to be true. If one could form one’s beliefs at will, then one could decide to believe *p* for (say) practical reasons, and not because *p* seems true. But then one wouldn’t believe *p* because *p* seems true, which, given the constitutive nature of belief, is impossible.

Now, my crude, highly enthymematic presentation of Williams’ argument raises a number of obvious questions, many of which have been taken up in the subsequent literature but will be suppressed here.⁷⁹ There’s also the fact that Williams’ argument targets voluntarism about *belief* rather than voluntarism *guessing*. (Pace Holguín (2022), it is not obvious that these claims are the same.) But to the extent I understand what it means for *belief* to “aim at truth”, it seems clear as anything that guessing aims at truth. In guessing the answer to a question, one is trying to get things right; to think *p* is to think *p* is true. But it does not follow that guessing is involuntary: as we know, our guesses can break ties and can exhibit various degrees of agnosticism, in a manner mediated by the activity of the will. Moreover, we can explain the sense in which guessing “aims at truth” in terms of cogency: one can only guess *p* if one takes *p* to be no less likely to be true than any of the relevant alternatives.⁸⁰ So the mere fact that some propositional attitude Φ aims at truth gives us little reason to believe that one cannot voluntarily Φ that *p*. Guessing is a straightforward counterexample. We should draw the same lesson the bouletic side of things: the mere fact that desire aims at the desirable gives us little reason to believe that one cannot voluntarily desire that *p*.

Outside of these sorts of theoretical arguments against bouletic voluntarism, one might have

⁷⁹ See Boespflug & Jackson (2024, §3.2) for a helpful survey.

⁸⁰ Here the relevant alternatives are the propositions defined on the relevant alternative set *Q* that include the same number of answers to *Q* as *p* does.

the following, more down to earth source of hostility toward the thesis. It just doesn't *seem* like any of our desires are formed at will, and a natural explanation of this appearance is that desires cannot be formed at will (at least not by human beings, anyway). This sort of thought was channeled in the opening paragraphs of this paper and is, I assume, the kind of thing Schopenhauer had in mind.

Unsurprisingly, I believe the appearances here are misleading. The impression that one's desires are *never* formed at will is likely the byproduct of an impoverished diet of examples, a diet which can be improved by attending to the kinds of cases discussed in §§6-7. However, even if the appearances aren't misleading in the case of some *particular* individual—even if a certain person (maybe the reader) truly never forms their desires at will, and perhaps couldn't even if they tried—all this would show is that *that* person does not (and perhaps cannot) form their desires at will. But it would be at least mildly chauvinistic to assume that one's own individual limitations are in fact species-grade. In any event, this section will present a model of bouletic voluntarism that allows for such interpersonal variation in our capacities for voluntary desire formation.

No doubt there are other ways of resisting bouletic voluntarism. Perhaps one could try to argue that bouletic voluntarism is incompatible with the "interpretationist" thesis that our ur-desires are determined by the facts about our behavioral dispositions, in suitably regimented senses of both 'behavior' and 'disposition' (Lewis 1974, Stalnaker 1984, Williams 2019). How exactly this argument would go is less than fully clear to me. In any event, we'll return to issues concerning the connections between desire and action in §14. So for now I'll set such concerns aside.

Alternatively, one could try to argue that the evidence for bouletic voluntarism is better explained by the hypothesis that our ur-desires are determined "at random" (at least folk psychologically speaking): some sub-personal selection process just happens to single out one of the candidate cogent propositions, which thereby becomes the agent's ur-desire. However, as stressed in §4, this kind of argument threatens to prove too much. Just about every paradigm case of voluntarism—word choice in subvocalization, certain instances of mental imagery and attention, and, most saliently, guessing—could plausibly be accounted for in such terms. But that's no strike against their voluntariness.

Keeping this dialectical filter in mind, I am inclined to regard the "It's just random" diagnosis not as a competing hypothesis, necessarily, but instead as a theory of what's going on when we want at will. That said, I also don't want to prove too much here. I do *not* want to endorse the inference from "It's just random" to "It is the result of the activity of the will". In fact, I will appeal to randomness in my preferred account of the grounds of *non*-voluntary ur-desire (§11). For better or for worse, the question of whether we can form desires at will is not going to turn on whether, at some level of explanation, the desires in question can be fairly described as having been determined at random. Instead, it's simply going to turn on whether in the relevant cases the process of desire formation sufficiently resembles whatever goes on when subvocalize, imagine, guess, and so on.

With that, I turn to the actual positive argument in favor of bouletic voluntarism, which is that the facts about what we desire can—and indeed often do—really seem chosen.

10.2 The case for bouletic voluntarism

Consider again the main Buridan case of §6. Equidistant from you are two essentially indistinguishable boxes labeled ‘Left’ and ‘Right’, each with a \$20 bill inside. You’re asked which, if any, you want/prefer. How might you answer?⁸¹

Well, when I imagine myself in such a case, the relevant phenomenology seems to me essentially the same as the phenomenology of being asked how I think a fair coin will land when tossed. With respect to the coin toss, I see no reason to favor heads over tails, or vice versa. Instead, I can simply pick one of the two arbitrarily. I guess heads (say) not because I think heads is more likely than tails; I guess heads because I think heads is *no less* likely than tails, and because that’s all the reason I need to make that choice. I could, of course, “opt out” of having a particular guess, and instead think merely that it will land some way or other. But that too looks to be a matter of my choosing.⁸²

The same appears to be true of Buridan cases. I see no reason to prefer Left or to prefer Right. Instead, I simply pick one of the two arbitrarily. I prefer (say) Left not because I think Left is more desirable than Right; I prefer Left because I think Left is no less desirable than Right, and because that’s all the reason I need to make that choice. I could, of course, “opt out” of having a preference, and instead want merely that I get either Left or Right. But that too looks to be a matter of my choosing.

Put another way: seeing the options as equally desirable, I can simply make up my mind to prefer Left to Right. If you were to ask, ‘Why do you prefer Left to Right? Is there something that makes you think that’s the better of the two boxes?’, I could easily reply, ‘No, I prefer Left to Right because why not?’, or—what amounts to the same—‘No, that’s just how I made up my mind’. Likewise, if you were to ask, ‘What would it have taken for you to prefer Right instead?’, my answer needn’t be ‘Right would have had to have been worth more than Left’, for I can just as well reply, ‘I would have had to have picked differently’. The nearest counterfactual possibilities in which I prefer Right to Left need not be possibilities in which Right happens to be more desirable than Left; they can be counterfactual possibilities in which I simply make up my mind differently. After all, without the facts about desirability there to guide me, what other basis could I have for my preference?⁸³

That’s the case for voluntarism about (some) Buridan preferences. It is worth stressing that I’m not committed to the thesis that Buridan preferences are *always* formed voluntarily. Perhaps Buridan preferences can sometimes be discovered rather than created, as it were. I am also open to the idea that there are some people who are simply incapable of forming Buridan preferences, and who are thus (pathologically) involuntarily indifferent. But we’ll return to these matters later in §11. For now the important point is this: desire-formation in cases of equidesirability looks a lot like guess-formation in cases of equiprobability; since the latter often involves the activity of the will, we should think the same is true of the former.

⁸¹ In case it matters (it shouldn’t), suppose you’ve been asked to keep the answer to yourself. As stressed in §6, everything can happen *in foro interno*.

⁸² Or can be, anyway. If it *had* to be a matter of my choosing, then we’d soon face a threat of regress. More on this sort of thing soon.

⁸³ Likewise, I don’t choose to have a preference because I think it’s better to have a preference than not. Again, my having a preference needn’t be necessary for getting a box; we could imagine the case as one where I’ve already been assigned one of the two boxes, and what I’m considering is which one, if either, I would prefer it to have been.

Next consider the example that drives §7's argument for the possibility of asceticism. You've just been gifted \$1,000,000, and may now receive an additional \$1, no strings attached. You're asked whether you want/prefer the free \$1. How might you answer?

Once again the relevant phenomenology seems to be analogous to the phenomenology of being asked what I think the outcome of a coin biased 50.1% in favor of heads will be. Putting myself in these circumstances, it's not like I simply notice what my state of mind is (or at least not invariably). Neither 'Well of course I think it will land heads' nor 'Well of course I don't have a particular view on the matter' need spring to mind, as it were. Nor is it like I must see some consideration that favors being agnostic to having an opinion (or vice versa). Instead, in typical cases, it seems to be up to me to choose whether to be opinionated (guess that the coin will land heads) or instead agnostic (guessing merely that the coin will land some way or other). I can simply decide arbitrarily whether to have an opinion or not.

The same appears to be true of the \$1 giveaway. It's not like I simply notice what my state of mind in the context of the \$1 giveaway. Neither 'Well of course I have a preference for it' nor 'Well of course I don't care; it's just a dollar' need spring to mind, as it were. Nor is it like I must see some consideration that favors having a preference to being indifferent (or vice versa). Instead, in typical cases, it seems to be up to me to choose whether to be picky (wanting the free \$1 in particular) or instead to be ascetic (wanting merely that I get either \$0 or \$1). I can simply decide arbitrarily whether to have a preference or not.

Put another way: seeing the options as sufficiently close to being equidesirable, it's up to me to choose whether to be indifferent to the free dollar. If you were to ask, 'Why don't you want the free dollar? Is it not worth something to you?', 'I could easily reply, 'No, it's not that it's not worth anything to me; it's just that it's not worth enough to make a difference to what I want'—or, more succinctly: 'Sure, but who cares?'. Likewise, if you were to ask, 'What would it have taken for you to prefer having the free dollar?', my answer needn't be 'The dollar would have had to have been worth more to me', for I can just as well reply, 'I would have had to have made up my mind differently'. The nearest counterfactual possibilities in which I prefer having a free dollar to not needn't be possibilities in which I regard having a dollar as more desirable than I actually regard it; they can be counterfactual possibilities in which I simply decide to care about such small differences.

That's the case for voluntarism about (some instances of) asceticism. As before, I'm not insisting that one's levels of asceticism are always determined voluntarily. There may be situations, just like the \$1 giveaway, where one simply "finds" oneself having or lacking the preference for the dollar, without the will playing a role in determining which it is. Moreover, it also seems to be a basic psychological fact that once the "gap" in desirability between prospects is sufficiently large, the will no longer plays any role in determining one's level of pickiness.⁸⁴ For example: if the free \$1 is upgraded to a free \$1,000,000 (still no strings attached), I am just going to prefer the free \$1,000,000, and there won't even be the faintest impression that the existence of that preference is a matter of my choosing. This suggests that the fact that \$1 is only somewhat more desirable than \$0 has a non-trivial role to play in my capacity for voluntary asceticism (more on which in §11).

⁸⁴ Our official conception of desirability is not well-equipped to handle talk of desirability "gaps". We'll come back to this issue later. For now it's enough to interpret 'gap in desirability' as 'gap in financial value'.

But—and for the last time—the claim is not that we *always* get to decide whether to have preferences between options that we know to be unequally good; it's that we *can and often do* get to decide as much. Once again the analogy to guessing is instructive. The claim that

We can and sometimes do decide whether to be opinionated or agnostic about whether a coin that is slightly biased towards heads will land heads when flipped.

is *not* undermined by the observation that

We usually can't help but be opinionated about whether a coin that is *overwhelmingly* biased towards heads will land heads when flipped.

Having a non-trivial capacity to determine the strength of one's opinions at will is one thing; having total control is another.

Summing up: bouletic voluntarism is true because wanting can be voluntaristic both with respect to tie-breaking and to asceticism. We can and often do decide *what* to want, as when we consider options that are known to be equally desirable. And we can and often do decide *whether* to want, as when faced between two or more options that are almost but not quite equally desirable. *Pace* Schopenhauer, man can want as he wills.

This in turn gives us a partial answer to the question we've been grappling with since the end of §9: what are the grounds of ur-desire? In at least some cases, *S*'s ur-desires are determined by the activity of *S*'s will. But why limit ourselves to a partial answer? Why not think that for *any* *S* and alternative set *Q*, the facts about what *S* ur-desires_{*Q*}—and thus wants_{*Q*}—are determined by *S*'s choice among the cogent_{*S,Q*} options?

The next section will answer this question. It will explain why the will isn't (and seemingly could not be) the grounds of ur-desire in the general case. And it will close by briefly considering the question of the grounds of non-voluntary ur-desire.

11 The limits of the will

I see at least two reasons to think that our ur-desires are not always formed voluntarily, both of which have already been hinted at. The first is that we often *find* ourselves having certain desires, without there being any impression that we formed them voluntarily. These are cases in which our desires are (in a manner of speaking) discovered rather than created. The second is that we often *can't help* but have the desires we have, thanks to the underlying desirability facts. These are cases in which our desires are (in a manner of speaking) the byproducts of sensuous impulses. I'll treat these cases in turn.

11.1 Limit one: preformation

Sometimes I notice that I'm thirsty. And often when I notice that I'm thirsty, I notice that I have a certain desire: namely a desire to quench my thirst. Usually there seems to be nothing remotely voluntaristic going on with any of this. It is not as if I notice I am thirsty, realize that this makes the proposition that I drink some water more desirable than the proposition that I don't, and then on the basis of this realization decide that I'm going to have a discriminating desire for

water rather than an indiscriminating desire for whatever may come. Instead, what usually happens is that I just find myself wanting to drink some water.

Taking these appearances at face value, we have reason to believe that some of our desires are not arrived at voluntarily: they are discovered rather than created.⁸⁵ I'll call such desires *preformed*.

It is not difficult to find other examples in which, intuitively, the relevant desires would most naturally be described as preformed. In fact, I suspect the vast majority of our desires are like this. As it turns out, there are strong theoretical considerations in favor of this suspicion. Here I'll mention three.

First, 'want' is highly context-sensitive. Again, this is supposed to be part of what explains why one can speak truly in answering both 'Yes' and 'No' to a question like 'Do you want to pay your taxes?', without contradicting oneself or changing one's mind about the value of anything. If one's desires didn't typically come preformed, then each true interpretation of '*S* wants *p*' would in principle have to correspond to a distinct act of will. It would have to be that relative to the "all things considered" desirability ordering, one chooses to want to pay one's taxes, while relative to the "what's intrinsically appealing" desirability ordering, one chooses to want to fail to pay one's taxes. Likewise for each of the other ways of interpreting *p* and 'want' so that '*S* wants *p*' expresses a truth. But it does not seem at all plausible that a plurality of acts of will are lurking beneath the surface of the plurality of ways in which we talk about desire. So plausibly at least some (if not the vast majority) of these desire states are not arrived at voluntarily, but instead come preformed.

Second, higher-order desires raise similar issues. Suppose *S* faces a choice between two equidesirable options *A* and *B*. Then *S* could either have a preference_{A,B} between *A* and *B* (it doesn't matter in which direction) or be indifferent_{A,B} between *A* and *B*. But what about *those* options—that is: the options of having a preference_{A,B} between *A* and *B* and being indifferent_{A,B} between *A* and *B*? Those options form an alternative set as well: {P_{A,B}, I_{A,B}}.⁸⁶ And *S* could easily have second-order desires with respect to that alternative set. In particular, *S* could either:

- Have a preference_{{P_{A,B}, I_{A,B}}} between having a preference_{A,B} or being indifferent_{A,B}
- Be indifferent_{{P_{A,B}, I_{A,B}}} between having a preference_{A,B} or being indifferent_{A,B}

So it seems *S* would have to determine their (second-order) desires_{{P_{A,B}, I_{A,B}}} at will too. But then what about *S*'s (third-order) desires_{{P_{{P_{A,B}, I_{A,B}}, I_{{P_{A,B}, I_{A,B}}}}...? Regress looms. We avoid it by allowing at least some (if not the vast majority) of *S*'s higher-order desires to come preformed.}}

Third and finally: an important theme of §§6-7 concerned the limits of our knowledge about what other people want in cases where there is no barrier to our knowing the relative desirability of the options in question (for the people in question). In the case of asceticism, the thought

⁸⁵ It's not mandatory to take these appearances at face value. As stressed in §3, the operative notion of 'voluntary' is semi-technical, and is meant to be understood in such a way that something's having been done subconsciously is compatible with it having been done voluntarily. Word-choice in subvocalized speech is meant to be a paradigm case of this. However, I'm taking it as a given that one's *feeling of thirst* is not generated voluntarily. And as far as I can tell, there are plenty of cases where the genesis of one's desire to quench one's thirst looks a lot more like the genesis of one's feeling of thirst than the genesis of the particular words that occur in some bit of sub-vocalized speech (at least with respect to questions of will).

⁸⁶ 'P_{A,B}' for 'preference_{A,B} between *A* and *B*', 'I_{A,B}' for 'indifference_{A,B} between *A* and *B*'.

was this:⁸⁷ knowing that someone finds A more desirable than B may well be a foolproof way of knowing whether they want_{A,B} B in particular (it's enough to know that they don't). But knowing that someone finds A more desirable than B is *not* a foolproof way of knowing whether they want_{A,B} A in particular or are instead indifferent_{A,B} between A and B. The answer to that question depends on how picky they are—a matter which is not settled by considerations of desirability alone.

But—and here's the new point—knowing that someone finds A more desirable than B *does* seem to be enough to know the disjunctive claim that either they want_{A,B} A in particular or are indifferent_{A,B} between A and B. If, however, the facts about what one desires_{A,B} could *only* be determined by one's acts of will, then the inference from 'S finds A more desirable than B' to 'S either wants_{A,B} A or wants_{A,B} A or B' would be valid only if the inference from 'S finds A more desirable than B' to 'S has made up their mind about what they want_{A,B}' were valid. And I see no reason whatsoever to think this latter inference is even reliably true, let alone valid. So again we have reason to think that our desires often enough come preformed.

In light of both the intuitive and theoretical considerations, I am inclined to endorse the following simple hypothesis. For any S and Q, S has a "default" ur-desire_Q, whose content is *not* determined by the activity of S's will.⁸⁸ In some cases—perhaps the vast majority—the fact that S ur-desires_Q p will not be a voluntary choice of S's, since (i) p is the default ur-desire_Q for S and (ii) S will not have performed any relevant acts of will with respect to p and Q. But in other cases—the cases that make (ii) relevant—S will have decided to ur-desire_Q p. And this will either be because p was not S's default ur-desire_Q (i.e., S changed their ur-desire_Q at will) or because S "endorsed" their preformed ur-desire.⁸⁹ In such cases, we may say that S desires_Q at will. As a crude analogy: at any given moment there might be something you are thinking about. When there is, often it will not be of your own choosing; rather, you will simply have found yourself in the midst of a train of thought (Strawson 2003). But often enough you can, by an act of will, either change what we're thinking about (i.e., revise it) or continue to sustain the thought (i.e., endorse it). The claim is that the voluntary aspect of desire has a similar character.

11.2 Limit two: the coercion of sensuous impulses

Next there is the more complex issue of freedom. I'll focus on the example briefly mentioned above: the \$1,000,000 (rather than \$1) giveaway.

Speaking for myself, I'd be lying if I said 'I don't care' when asked whether I want a free \$1,000,000. Moreover, it is a psychological fact that there's nothing I can really do to change this about myself. My desire for the \$1,000,000 is not merely preformed; it is immutable. At first blush, this all seems to have something to do with the fact that I find the prospect of getting a free \$1,000,000 *considerably* more desirable than the prospect of getting nothing. *Mutatis mutandis*

⁸⁷ The same points go through (*mutatis mutandis*) for tie-breaking.

⁸⁸ We could weaken this view to require that S stand in some kind of "comprehension" relation to p and/or Q in order to count as ur-desiring_Q anything at all. But I'll stick with the simpler view for concreteness.

⁸⁹ I don't intend for this talk of "endorsement" to be particularly illuminating. It's just a label for whatever is going on in cases where an agent makes up their mind to ur-desire_Q p even when p is their default ur-desire_Q. And we should expect this sort of thing to happen—and so should want to have a label for it—since: (i) we have strong abductive reasons to think that agents can have desires_Q whether or not they make up their mind to; (ii) agents can nonetheless make up their mind to have particular desires_Q; and (iii) there is no reason to expect the desires_Q they ultimately make up their mind to adopt to differ from the ones they had before doing so.

for my desires relative to $\{\$0, \$1,000,000, \$1,000,001\}$: perhaps I can choose whether I want the $\$1,000,001$ in particular or am instead indifferent between it and the $\$1,000,000$ (what's the extra dollar if I'm already getting a million?). But I cannot help but desire that I get at least one of the two big payouts, since they're both just so much more desirable than the remaining alternative.

Examples like this suggest that certain of our desires—perhaps even the vast majority of them—are such that we couldn't have easily failed to have them.⁹⁰ I submit that such desires are *not* formed voluntarily. This is not because *being voluntary* necessarily entails *could easily do otherwise*. As is well known, the relationship between these two properties is a complex matter.⁹¹ The point is just that certain of our desires—say, my desire for $\$1,000,000$ (rather than $\$0$)—seem no more deserving of the label 'chosen' than do the laws of physics, the circumstances of our births, or the contents of our intrusive thoughts.

Thus, independently of the issue of preformation, there is reason to believe that a significant portion of our desires are not in any sense determined by our acts of will. Sometimes the facts about desirability are too overpowering. In light of this, how should we be thinking of the voluntary aspect of desire?

Start with a label. Say that for any agent S and alternative set Q , there is some non-zero number of propositions defined on Q that are *willable* for S relative to Q (' $\text{willable}_{S,Q}$ '). Here are the three main assumptions I'll be making about willability:

1. I will assume that, necessarily, the $\text{willable}_{S,Q}$ propositions are a non-empty subset of the $\text{cogent}_{S,Q}$ propositions.
2. I will assume that (i), necessarily, whenever there is more than one $\text{willable}_{S,Q}$ proposition, S could easily form their ur-desire_Q at will, by choosing among the $\text{willable}_{S,Q}$ propositions; and (ii), necessarily, whenever there is exactly one $\text{willable}_{S,Q}$ proposition, S involuntarily ur-desires_Q that proposition, and could not have easily done otherwise. That is to say: I am assuming that the existence of more than one $\text{willable}_{S,Q}$ proposition is necessary and sufficient for S to be in a position to form their ur-desires_Q voluntarily.⁹²
3. I will assume that, necessarily, whether a proposition p is $\text{willable}_{S,Q}$ is determined by two things:
 - (a) S 's desirability ordering.
 - (b) Certain contingent features of S 's psychological constitution (more on which in a moment).

So although I am assuming that it is a matter of metaphysical necessity that S 's ur-desire_Q is $\text{willable}_{S,Q}$, I am *not* assuming that the relationship between desirability (for S) and $\text{willability}_{S,Q}$ is itself metaphysically necessary. The relationship between desirability (for S) and what's $\text{willable}_{S,Q}$ can vary across worlds (and times), in accordance with the relevant ways in which S 's psychological constitution can vary.

⁹⁰ (At least holding fixed the relevant aspects of the circumstances in which they were formed, etc.)

⁹¹ See, e.g., Ginet (1966), Frankfurt (1969), van Inwagen (1983), Fischer (1994), Pickard (2013), Nichols (2015).

⁹² Note that if we were to set aside issues related to preformation, we could upgrade from the hypothetical 'be in a position to form' to the categorical 'form'.

So, our powers of voluntary desire formation are determined by considerations of willability_{S,Q}, which are in turn determined by some function of the desirability of the options in *Q* and certain (as-of-yet unspecified) features of *S*'s psychological constitution. What might this function look like?

Well, as I've tried to emphasize, there is considerable flexibility in how discriminating our desires are, even in cases where the relevant desirability facts are comparatively inflexible. Some people seem to find it easy to pick discriminating desires in Buridan cases; others seem not to. Likewise, some people seem to find it easy to choose to be ascetic in the face of small differences; others seem not to. And in the other direction, some people cannot help but be indifferent to small differences—forced indifference in Buridan cases being a limiting case of this—while other people seem to have no problem with forming preferences at will in such cases. So even restricting attention to ordinary human beings, the relationship between desirability and willability is bound to be somewhat complex, and to vary from person to person.

By way of making progress on this issue, start by noting that there are two ways in which one's desires could be forced. It could be that one cannot help but have a desire that discriminates between some number of the options in question. I'll call this a "forced preference". But it could also be that one cannot help but have a desire that *fails* to discriminate between some number of the options in question. I'll call this a "forced indifference" (for lack of a better term). Someone who can't help but want \$1,000,000 rather than \$0 has a forced preference; someone who can't help but be indifferent between two options in a Buridan case has a forced indifference.

A natural hypothesis, which I will be adopting without argument for the remainder of this section, is that cases of forced preference are due to certain options being *sufficiently less* desirable than certain other options in question, while cases of forced indifference are due to certain options being *sufficiently close* to equidesirable with certain other options in question.

Somewhat more precisely: suppose *q* is the most desirable option in *Q* (or tied for most desirable) and that *p* is some other option in *Q*. When the gap in desirability between *p* and *q* is sufficiently large, the only willable_{S,Q} propositions will be those that entail $\neg p$ (since *p* is so much worse than the best option on the table, *q*). And when the gap between *p* and *q* is sufficiently small, the only willable_{S,Q} propositions will be those that are entailed by both *q* and *p* (since *p* is so close to *q*). Of course, since different people exhibit different powers of voluntary desire formation—and since there is no reason to expect that one's actual powers of voluntary desire formation are part of one's essence—the facts about which kinds of differences are "sufficiently large" or "sufficiently small" will be subject-sensitive and contingent. But these facts, in combination with the facts about desirability, exactly settle the scope of one's voluntary powers of desire formations (at a given world and time).

That's the kind of story I want to explore. But first an immediate difficulty must be addressed. If we want to be able to say things like 'The reason *S* cannot help but want the \$1,000,000 is that \$1,000,000 is so much more desirable than \$0', then we better be able to talk meaningfully about *degrees* of desirability. But given our austere assumptions about the workings of desirability orderings—assumptions that allow for incomparabilities and intransitivities—nothing in the framework guarantees that we can sensibly talk this way. So if we want to go beyond a purely boilerplate conception of *willability*, we'll need desirability orderings to have enough structure

to allow for cardinal (rather than merely ordinal) comparisons of desirability. Thus, for the remainder of this section I will work under the assumption (disavowed in §9.1) that desirability orderings have the structure they need to be represented by (expected) utility functions. Whereas before we were limited to talking of options as being less, more, or equally desirable, now we'll be able to talk about just *how* desirable they are, whether they have positive or negative (or neutral) desirability, and so on. This extra structure will let us propose and evaluate a number of substantive hypothesis about the limits of our powers of voluntary desire formation—hypotheses that would be simply unintelligible relative to the logically austere conception of desirability we've been employing so far.

With that, my hypothesis is that every agent S is associated with two “thresholds of willability”: (i) Δ_S , a threshold of forced preference; and (ii) δ_S , a threshold of forced indifference. Though the particular values of these thresholds can change across worlds and times, they always satisfy the following basic constraints:

- Necessarily, $0 \leq \Delta_S \leq \infty$. That is: Δ_S is either a non-negative real number or larger than every real number.
- Necessarily, $-\frac{1}{\infty} \leq \delta_S \leq \infty$. That is: δ_S is either an arbitrarily small negative real number, or a non-negative real number, or larger than every positive real number.
- Necessarily, $\Delta_S \geq \delta_S$.

Presently these claims should seem a bit opaque (especially the stuff involving “ ∞ ”); their meaning and significance will be easier to grasp once they've been put to use.

Start with forced preference. Let ‘ $\max_{S,Q}$ ’ be shorthand for ‘the expected utility of the highest expected utility option(s) in Q for S ’.⁹³ Then we can say:

SUFFICIENT DIFFERENCE

For any Q and any $p \in Q$, if $\max_{S,Q} - EU_S(p) > \Delta_S$, then: for any q , q is willable $_{S,Q}$ only if q is *not* entailed by p .

What SUFFICIENT DIFFERENCE says, effectively, is that whenever the difference in the expected utility between the best option and p exceeds Δ_S , one cannot help but be such that p is incompatible with getting what one *ur-desires*. To see how this works, let's assume that $\Delta_S = EU_S(\$5)$. Then, relative to $\{\$0, \$6\}$, SUFFICIENT PROXIMITY says that S cannot help but want \$6. And relative to $\{\$0, \$5, \$10\}$ it says that S can't help but want *not* to get \$0. But since the gap between \$5 and \$10 does not exceed Δ_S , SUFFICIENT DIFFERENCE is silent on whether S is maximally picky—wanting $_{\{\$0, \$5, \$10\}}$ \$10 in particular—or partially ascetic—merely wanting $_{\{\$0, \$5, \$10\}}$ either \$5 or \$10.

More generally, when $\Delta_S = 0$, one cannot help but be maximally picky—always preferring better options to worse, no matter how small the difference. One's preferences are pathologically

⁹³ If Q contains infinitely many unboundedly-improving alternatives, then $\max_{S,Q}$ will be undefined. For example: suppose a coin is going to be flipped in some super-task fashion until it lands tails. For each $n \geq 0$, if the coin lands heads exactly n times before coming up tails, you'll get to spend 2^n days in heaven. How many times do you want the coin to land heads? Supposing that ‘infinitely many’ is not an admissible answer, I don't really know what you're supposed to say. So for the purposes of this paper I'm just going to ignore such cases and assume that $\max_{S,Q}$ is always defined.

necessitated, to borrow a phrase of Kant's. However, as Δ_S increases, so does one's capacity for voluntary asceticism (at least as far as SUFFICIENT DIFFERENCE is concerned). Supposing it is possible for there to be agents who *never* have forced preferences, Δ_S needs to be able to take on values that exceed the expected utility of every proposition. In such cases we say that $\Delta_S = \infty$.⁹⁴

Next there is forced indifference:

SUFFICIENT PROXIMITY

For any Q and any $p \in Q$, if $\max_{S,Q} - EU_S(p) \leq \delta_S$, then: for any q , q is willable _{S,Q} only if q is entailed by p .

What SUFFICIENT PROXIMITY says, effectively, is that whenever the difference in the expected utility between the best option and p fails to exceed δ_S , one cannot help but be such that p is a way of getting what one ur-desires. So supposing $\delta_S = EU_S(\$5)$, SUFFICIENT PROXIMITY says that S cannot help but be indifferent _{$\{\$0,\$5\}$} between \$0 and \$5; nor can S help but be indifferent _{$\{\$5,\$10\}$} between \$5 and \$10. But as far as SUFFICIENT PROXIMITY is concerned, S could well fail to be indifferent _{$\{\$0,\$10\}$} between \$0 and \$10 (by desiring _{$\{\$0,\$10\}$} \$10 in particular). What about relative to $\{\$0, \$5, \$10\}$? Here SUFFICIENT PROXIMITY says that S cannot desire _{$\{\$0,\$5,\$10\}$} \$10 in particular—since \$5 is too close to it—but is otherwise silent on whether S has the partially ascetic desire _{$\{\$0,\$5,\$10\}$} for $\$5 \vee \10 or the fully ascetic desire _{$\{\$0,\$5,\$10\}$} for $\$0 \vee \$5 \vee \$10$. As far as SUFFICIENT PROXIMITY is concerned, that's up to S .⁹⁵

More generally, as δ_S decreases, one's capacity for voluntary pickiness increases: when $\delta_S = 0$, indifference is forced only in Buridan cases; when $\delta_S = -\frac{1}{\infty}$, indifference is never forced.⁹⁶ But in the case where $\delta_S = \infty$ (should there be one), indifference is pathologically necessitated.

SUFFICIENT DIFFERENCE and SUFFICIENT PROXIMITY impose necessary (though not sufficient) conditions on willability. SUFFICIENT DIFFERENCE says that a proposition is willable only if the alternatives it includes are not more than Δ_S from the best alternative. SUFFICIENT PROXIMITY says that a proposition is willable only if it includes every alternative that is within δ_S of the best alternative. These principles are consistently combinable only if $\Delta_S \geq \delta_S$ —hence why we're assuming that as well. Otherwise both principles are compatible with the full range of capacities for voluntary desire formation, whether among actual or merely possible agents. Making SUFFICIENT DIFFERENCE compatible with the possibility of maximal asceticism is just a matter of finding worlds where, for some agent S , $\Delta_S = \infty$. And making SUFFICIENT PROXIMITY compat-

⁹⁴ Note that ' ∞ ' is not meant to denote a *number*. All that matters is that Δ_S can take a value such that for any p , it's not the case that $\max_{S,Q} - EU_S(p) > \Delta_S$. Calling this value ' ∞ ' just makes it easier to think about what's going on intuitively: no difference between options is so large as to force a preference between them. Similar points will apply to δ_S and $-\frac{1}{\infty}$ below.

⁹⁵ It is for exactly this reason that SUFFICIENT PROXIMITY is concerned with the desirability gap between p and the best option in Q , rather than between p and any other option in Q . Asceticism can come in degrees, so indifference is non-transitive (cf. §7). And because indifference is non-transitive, it can't be that, relative to every Q , any two options that are within δ_S of one another are such that every willable _{S,Q} proposition is entailed either by both of them or by neither of them. Consider, e.g., $Q = \{\$0, \$5, \$10, \$15\}$ when $\delta_S = EU_S(\$5)$. If it were the case that *any* two options that are within δ_S must be such that every willable _{S,Q} proposition is either entailed by both or neither of them, then the only willable _{S,Q} proposition would be the maximally ascetic $\$0 \vee \$5 \vee \$10 \vee \15 . By contrast, as far as SUFFICIENT PROXIMITY is concerned, indifference is forced only with respect to the propositions that are within δ_S of $\max_{S,Q}$.

⁹⁶ Note that like ' ∞ ', ' $-\frac{1}{\infty}$ ' is also not meant to denote a number. All that matters is that δ_S can take a value such that for any p , it's not the case that $\max_{S,Q} - EU_S(p) \leq \delta_S$. Calling this value ' $-\frac{1}{\infty}$ ' makes it easier to think about what's going on intuitively: no difference between options—not even zero difference—is too small to force indifference between them.

ible with the possibility of maximal pickiness is just a matter of finding worlds where, for some agent S , $\delta_S = -\frac{1}{\infty}$.

The conjunction of SUFFICIENT DIFFERENCE and SUFFICIENT PROXIMITY implies that whenever $\Delta_S = \delta_S$, S 's desires are completely outside of S 's control. This is because when $\Delta_S = \delta_S$, the propositions that SUFFICIENT DIFFERENCE says S is free to be indifferent between are propositions that SUFFICIENT PROXIMITY says S is forced to be indifferent between, while the propositions that SUFFICIENT PROXIMITY says S is free to have a preference between are propositions that SUFFICIENT DIFFERENCE says S is forced to have a preference between.⁹⁷ Thus, in the context of this framework, Schopenhauer's dictum can be reformulated as follows:

SCHOPENHAUER'S DICTUM

For ordinary human beings, it is always the case that $\Delta_S = \delta_S$.

And the "what's best" theory of desire—a theory that abides by Schopenhauer's dictum—is equivalent to the thesis that:

WHAT'S BEST

Necessarily, for every S : $\Delta_S = \delta_S = 0$.

For if $\Delta_S = \delta_S = 0$, then S cannot help but desire _{Q} the unique best option in Q (if there is a unique best option) or the disjunction of all the tied-for-best options (if there isn't). I have argued, by contrast, that it's not even contingently true that for every human being S , $\Delta_S = \delta_S = 0$.

Things get more interesting if we assume that the *only* substantive conditions on willability _{S,Q} are those given by the conjunction of SUFFICIENT DIFFERENCE and SUFFICIENT PROXIMITY. So let's make that assumption. Let's assume, that is, that of the various cogent _{S,Q} propositions, the ones that are willable _{S,Q} are all and only those that are consistent with both SUFFICIENT DIFFERENCE and SUFFICIENT PROXIMITY, given the relevant values of Δ_S and δ_S .⁹⁸ It then follows that the only conditions under which your desires are completely out of your control are those in which $\Delta_S = \delta_S$. This in turn lets us reformulate bouletic voluntarism as follows:

BOULETIC VOLUNTARISM

For ordinary human beings, it can be and often is the case that $\Delta_S \neq \delta_S$.

For whenever $\Delta_S \neq \delta_S$ —whenever there is a gap between the range of forced indifference and the range of forced preference—there will in principle be room for the exercise of the will. For example: if $\Delta_S = 0$ and $\delta_S = -\frac{1}{\infty}$, then although S cannot be an ascetic, it is within S 's

⁹⁷ In detail: Consider any $Q = \{p_1, \dots, p_n\}$. Suppose p_1 is the top-ranked propositions (it makes no difference if multiple propositions are tied for first). Given SUFFICIENT DIFFERENCE, for any $p \in Q$, if $EU_S(p_1) - EU_S(p) > \Delta_S$, then S 's ur-desire _{Q} is forced to be not entailed by p . And given SUFFICIENT PROXIMITY, for any $p \in Q$, if $EU_S(p_1) - EU_S(p) \leq \delta_S$, then S 's ur-desire _{Q} is forced to be entailed by p . But supposing $\Delta_S = \delta_S$, it follows that for any $p \in Q$, either $EU_S(p_1) - EU_S(p) > \Delta_S$ or $EU_S(p_1) - EU_S(p) \leq \delta_S$. (Recall that, necessarily, $\Delta_S \geq 0$; so if $\Delta_S = \delta_S$, then $\delta_S \geq 0$.) So for any p in Q , either S 's ur-desire _{Q} is forced to be entailed by p or S 's ur-desire _{Q} is forced to be not entailed by p . So there is only one willable _{S,Q} proposition— S 's ur-desire _{Q} as a whole is forced.

⁹⁸ This hypothesis is entailed by our background assumptions about willability plus the following two further assumptions:

Δ -DEPENDENCE For any Q and $p \in Q$, if every willable _{S,Q} proposition entails $\neg p$, then: $\max_{S,Q} - EU_S(p) > \Delta_S$.

δ -DEPENDENCE For any Q and $p \in Q$, if every willable _{S,Q} proposition is entailed by p , then: $\max_{S,Q} - EU_S(p) \leq \delta_S$.

powers to voluntarily form Buridan preferences. And if $\Delta_S > 0$ and $\delta_S = 0$, then although S cannot form Buridan preferences, it is within S 's powers to voluntarily be indifferent to at least some difference. More generally, we can think of the gap between Δ_S and δ_S as S 's "band of freedom": the larger the gap, the greater the scope of S 's voluntary powers. At one extreme, $\Delta_S = \delta_S$, the gap is zero, and S 's ur-desires are invariably forced. At the other extreme, $\Delta_S = \infty$ and $\delta_S = -\frac{1}{\infty}$, the gap is unbounded, and it is within S 's powers to pick any cogent ur-desire.

As a matter of empirical fact, no human being exercises full voluntary control over their desires. It is just a fact of life that there are things we cannot help but want. In our present framework, this is tantamount to the claim that every human being has a finite Δ_S . Indeed, a typical human being's Δ_S seems to be quite small—on the order of the disvalue of a mild headache, or the value of a few dollars. This why all the examples in §7 concerned putative cases of indifference to *small* differences. But if my claims in that section are correct, then it's relatively common for Δ_S to at least be non-zero. And this in turn is why it is usually not so hard to find propositions that are close enough to being equally desirable that one is free to choose between preference and indifference. Of course, some people (and perhaps most animals) may well be incapable of this; for them, $\Delta_S = 0$.

In the other direction, it seems less common for human beings to be *incapable* of forming discriminating desires in cases of known difference. This is just to say that involuntary asceticism seems rarer than involuntary pickiness. And this suggests that δ_S is rarely much greater than 0, at least in typical cases. But perhaps certain kinds of *anhedonia* can be characterized in terms of abnormally high δ_S values (cf. [Stocker 1979](#) and footnote 43 above). That said, some people report an inability to form discriminating desires in Buridan cases. Supposing these reports are accurate, these are people for whom $\delta_S \geq 0$. But I suspect that most humans can and often do tie-break at will, and are thus such that $\delta_S = -\frac{1}{\infty}$.

Can we say anything about what determines the value of Δ_S and δ_S , for some particular agent S ? There are a few options here, which I'll mention only very briefly. We could, as always, take the facts about Δ_S and δ_S to be folk-psychologically fundamental. Alternatively, we could try to give a reductive account in modal terms: perhaps the facts about Δ_S and δ_S are determined by the facts about which ur-desires S could easily have formed (holding fixed their desirability ordering and so on). Or we could try to give a reductive account in terms of some other psychologically significant notion, like *willpower*.

In fact, this last proposal is the one I find most appealing, though developing it in full detail would take us too far afield. So for now I'll just note that the relevant notion of 'willpower' isn't meant to be the standard philosophical one—which, very roughly, is about one's capacity to *resist temptation* ([Holton 2009](#), Chs.5-6, [Sripada 2010, 2012](#)). It's not like when presented with the prospect of a free \$1,000,000 I invariably succumb to a second-order desire to have a first-order preference for it. I'm going to have that first-order preference whether or not there is any mismatch between my higher-order desires or prior intentions. As such, the relevant notion of 'willpower' seems to me better understood in terms of the kind of "psychic powers" involved in our capacities for imagination and attention (cf. [James 1890](#)). My inability to fail to desire the \$1,000,000 is less like my inability to keep my promises, and more like my inability to bring certain (sufficiently detailed) images to mind, or my inability to stop thinking about pink

elephants once I've started. In *this* sense of 'will', I lack the force of will to resist the coercion of (sufficiently) sensuous impulses. But my meager force of will is just a contingent feature of my psychological constitution. It is not metaphysically impossible that I acquire the capacity to be voluntarily indifferent to the prospect of a free \$1,000,000.

That covers most of what I want to say about coercion of sensuous impulses. But before moving on to this section's final topic, I want to briefly amplify one point.

What I have presented in this subsection is intended to be taken as a *model* of the manner in which the facts about desirability interact with the voluntary aspect of desire (over and above considerations of cogency). I do not present it under the pretense that it might be a final theory of the phenomenon. Indeed, my particular model—stated in terms of Δ_S and δ_S —has a number of obvious limitations. It requires the ambitious psychological assumption that an agent's desirability ordering can be represented by an expected utility function. It also assumes that the coercive force of sensuous impulses is always a matter of the *distance* in the desirability of the options in question. But it is plausible that considerations of absolute desirability and/or ratios of desirability could sometimes be relevant too. Perhaps some agents can't help but be maximally picky when considering options that have "negative" desirability, even if they can exercise voluntary indifference with respect to options that have "positive" desirability. Or perhaps some agents' eyes start to glaze over when considering options whose expected utility exceeds some sufficiently large threshold, at which point the ratios matter more than the gaps. The model also assumes that the values of Δ_S and δ_S are subject-matter independent. But this too might be unrealistic. Being able to freely choose one's preferences in a standard Buridan case is one thing; being able to freely choose one's preferences in a scenario like that of Sophie's choice is quite another.

Accounting for these subtleties is a task for another occasion. For now I will press on with it as we turn to this section's final topic, which concerns the grounds of *non-voluntary* desire.

11.3 The grounds of non-voluntary desire

As we saw above, if $\Delta_S = \delta_S$ (and the facts about willability are exhausted by SUFFICIENT DIFFERENCE and SUFFICIENT PROXIMITY), then *all* of S 's desires are formed involuntarily, in accordance with the facts about S 's desirability ordering and the values of Δ_S and δ_S . For such agents, the question of what determines their non-voluntary desires reduces to the question of what determines the values of Δ_S and δ_S . And I will suppose that we're happy to end our search for an explanation there.

But what if S , like many of us, has a non-zero band of freedom—i.e., $\Delta_S - \delta_S > 0$? Then there will be some alternative set $Q = \{p, q\}$, where $q >_S p$, such that: it can be a matter of S 's choosing whether S 's ur-desire $_Q$ includes only q or instead includes both p and q . For example, if $\Delta_S = \$5$ and $\delta_S = \$0$, and $Q = \{\$1, \$5\}$, then S 's ur-desire $_{\{\$1, \$5\}}$ could easily be $\$5$ or could easily be $\$1 \vee \5 . Which of the two it ends up being will, in at least some cases, be determined by S 's acts of will.

However, the issue of preformation gives us good reason to think that *prior* to S performing any relevant acts of will, there is a fact of the matter as to which of $\$5$ or $\$1 \vee \5 is the content of S 's ur-desire $_{\{\$1, \$5\}}$. The question is what determines which it is. Obviously the facts about S 's

desirability ordering and the values of Δ_S and δ_S won't do it, since the gap between \$1 and \$5 falls within S 's band of freedom (i.e., $\delta_S < [EU_S(\$5) - EU_S(\$1)] \leq \Delta_S$). So then what does?

Here is the model I favor. Suppose that in addition to Δ_S and δ_S , each agent is also associated with a “critical point” $\heartsuit_S$, whose value is contingent but guaranteed to abide by the following two constraints:

- Necessarily, $\Delta_S \geq \heartsuit_S \geq \delta_S$.
- Necessarily, $\heartsuit_S \geq 0$.

We can then use $\heartsuit_S$ to determine S 's non-voluntary ur-desires as follows:

NON-VOLUNTARY UR-DESIRE

S non-voluntarily ur-desires _{Q} p iff:

- (i) p is defined on Q .
- (ii) For each $q \in Q$ such that $\max_{S,Q} - EU_S(q) \leq \heartsuit_S$: q entails p .
- (iii) There is no p^* defined on Q such that S has chosen to ur-desire p^* .

In words: if S hasn't chosen their ur-desire _{Q} at will, then S non-voluntarily ur-desires _{Q} the disjunction of all the propositions in Q that are within $\heartsuit_S$ of the best option in Q .⁹⁹

A few comments on this analysis are in order.

First, one nice feature of this sort of analysis is that it predicts that one's patterns of non-voluntary (ur-)desire formation are reasonably *regular*. If, say, $\heartsuit_S = EU_S(\$1)$, then we know that S will default to indifference between options that are within a dollar's worth of desirability of the best option on the table, and will default to dispreference for options that are more than a dollar's worth of desirability away from the best option on the table. The sub-personal processes responsible for generating S 's non-voluntary desires needn't decide each question afresh, as it were.

Second, this analysis, as stated, predicts that it is impossible to tie-break non-voluntarily. That is to say: if NON-VOLUNTARY UR-DESIRE is correct, then the only way for S to have a discriminating desire in a Buridan case is for S to have *chosen* to have such a desire. This is due to the stipulation that $\heartsuit_S$ is always non-negative. We could abandon this stipulation if we wanted, allowing $\heartsuit_S$ to take on the kind of negative infinitesimal values δ_S can sometimes have. If we were to do that, however, we'd have to complicate clause (ii) of the analysis. We'd essentially have to make it disjunctive, treating separately two cases: (a) the case in which $\heartsuit_S \geq 0$, which can be left as is. And (b) the case where $\heartsuit_S < 0$, for which we'd have to allow S 's non-voluntary ur-desire _{Q} to either be one of the tied-for-best options in Q , or some disjunction of the tied-for-best options—with the matter of which of *those* it is presumably being determined at random. In addition to being aesthetically unpleasing, this modified analysis would lose some of the “regularity” the original version provides. Consequently, I'm inclined to stick with NON-VOLUNTARY UR-DESIRE as it is, and to treat its “no non-voluntary tie-breaking” consequence as a substantive prediction worth taking seriously.

⁹⁹ Apropos footnote 70 above, we might want to add a clause requiring S to have some bouletic attitude or other with respect to Q , to account for the possibility of cases in which the options in Q are beyond S 's grasp, or what have you. But I'll continue to suppress such complications.

Third and finally, there is a question of the grounds of $\heartsuit_S$. After all: unless we have a sense for what could determine its value, it's not clear to what extent we've cleared anything up with respect to the matter of the grounds of non-voluntary ur-desire; it feels more like we've just slapped a label onto the phenomenon. Now, if S 's band of freedom is empty, the value of $\heartsuit_S$ is settled by the value(s) of Δ_S and δ_S . And so in that case at least we're not saddled with any extra explanatory baggage. But of course not every agent is so thoroughly hostage to the coercion of sensuous impulses. If S can want at will, then Δ_S and δ_S constrain but do not determine what S non-voluntarily desires, leaving a range of possible values for $\heartsuit_S$.

So what determines the value of $\heartsuit_S$ in the general case? Here I'll briefly mention two options. One option is to embrace the brute fact view. There is nothing folk-psychologically speaking that settles the value of $\heartsuit_S$; it is just a random point somewhere between Δ_S and δ_S . Understood as such, the $\heartsuit_S$ model does indeed constitute a kind of explaining by labeling. Maybe that's all there is to say. But the model at least has the virtue of reducing a broad class of unexplained facts—e.g., S 's non-voluntary ur-desire $_{Q_1}$, S 's non-voluntary ur-desire $_{Q_2}$, ...—to facts about S 's desirability ordering and the value of $\heartsuit_S$.

Another (at least superficially distinct) option is to go for some kind of modal analysis. On one version—a kind of counterfactual analysis—the facts about $\heartsuit_S$ are determined by the facts about what the value(s) of Δ_S and δ_S would have been, had S 's band of freedom been empty (holding fixed S 's desirability ordering and so on). On another version of the modal analysis, the facts about $\heartsuit_S$ are determined by S 's dispositions to form non-voluntary desires. Why is it the case that $\heartsuit_S = EU_S(\$1)$? Well, because S is disposed to be indifferent to \$1 differences and to have preferences for greater-than-\$1 differences in situations where their desires are not chosen (and their desirability ordering is held fixed, and so on). No doubt there are many other kinds of modal analyses available.¹⁰⁰ That said, my sense is that any plausible version of a modal analysis will require such a fine-tuned interpretation of the relevant modal operators so as to make the proposal a version of the brute fact view in all but name. However, getting to the bottom of such matters is a task for another occasion.

Taking stock

Here is a way of seeing how the pieces developed over the past few subsections fit together. Suppose we've got some set of options Q . Which of the propositions defined on Q could, meta-physically speaking, be the object of S 's ur-desire $_Q$? Well, in principle any of them: we just need to find a world in which S 's desirability ordering and thresholds of preference (Δ_S) and indifference (δ_S) have the right values. But if we restrict the domain of the 'could' to worlds in which S 's desirability ordering is as it actually is, our answer becomes more constrained: now the only propositions (defined on Q) that could be the object of S 's ur-desire $_Q$ are the cogent $_{S,Q}$ ones. But in principle it could still be any cogent $_{S,Q}$ proposition, since again we can always find worlds in which Δ_S and δ_S have the right values (a trivial case: $\Delta_S = \infty$ and $\delta_S = \frac{1}{\infty}$). So

¹⁰⁰ Indeed, one could eschew the whole apparatus of ur-desire in terms of $\heartsuit_S$ and instead try to ground the facts about non-voluntary ur-desire in terms of behavioral dispositions (something like: one's ur-desires $_Q$ are determined by the actions one is disposed to take when faced with a choice between the options in Q). We'll be returning to the connections between desire and action in some detail in §14. Here I'll just assert without explanation that the prospects of such a reduction strike me as unpromising.

let's restrict the domain of the 'could' by that too: that is, let's only look at worlds that agree with actuality on the value of Δ_S and δ_S . Now the only propositions that could be the object of S 's ur-desire_Q are the $\text{willable}_{S,Q}$ ones. If there's only one $\text{willable}_{S,Q}$ proposition, then we've ended our search: that's the only proposition defined on Q that could be the object of S 's ur-desire_Q . However, if there's more than one $\text{willable}_{S,Q}$ proposition—which let's suppose there in fact is—then any of the $\text{willable}_{S,Q}$ propositions could be the object of S 's ur-desire_Q . We just need to find worlds where S performs the relevant acts of will. But what if we restrict our attention to worlds where S *hasn't* performed any relevant acts of will. Now which propositions could be the objects of S 's ur-desire_Q ? The answer remains: any of the $\text{willable}_{S,Q}$ propositions. For now we just need to find worlds where $\heartsuit_S$ has the right value. (And we know that if there is more than one $\text{willable}_{S,Q}$ proposition, then $\Delta_S > \delta_S$, and so $\heartsuit_S$ has room to vary even across worlds that hold Δ_S and δ_S fixed.) So what if we hold $\heartsuit_S$ fixed too? Only then, finally, will we have landed upon a unique proposition. This is guaranteed to be S 's actual ur-desire_Q —at least until S chooses to change it.

This leaves us with the following general picture of the voluntary aspect of desire. Our desires are a function of our ur-desires , which are themselves constrained but not determined by considerations of desirability. Desirability constrains ur-desire via cogency. But it does not determine ur-desire , since for any possible alternative set there will invariably be more than one cogent proposition, and thus more than one candidate ur-desire . So what determines our ur-desires ? Sometimes it is our acts of will: we make up our minds to ur-desire what we do. But sometimes the will is silent. Sometimes we ur-desire what we do simply because that ur-desire came preformed, and we failed to overrule it or endorse it; other times we ur-desire what we do because we are pathologically necessitated—the values of Δ_S and δ_S are too close together to afford any latitude in our ur-desires . But the extent of the pathology is a contingent matter. Though no actually existing human being is capable of exerting total control over their desires, this doesn't entail that we *couldn't have had* such a capacity, or—perhaps more conservatively—that there couldn't have been agents who have it, whether human or not.

The final three sections of the paper will cover some loose ends, as well as some of the theoretical implications of the theses I've defended so far. §12 will speak to the question of the grounds of desirability. §13 will take up the topic of rational desire. §14, the final section of the paper, will focus on the role of desire in action, both descriptively and normatively. I cannot hope to attend to any of these topics in the detail it deserves. The aim of these closing sections of the paper is more to get a bird's-eye view of how bouletic voluntarism and the cogency-theoretic account of desire interact with some longstanding questions about the nature and norms of desire.

12 Desirability redux

Up until now we've been treating desirability as a theoretical primitive. It's the kind of value relevant to desire, whatever that kind of value may be. Is there any hope of a reductive analysis of the notion?

Suppose for the moment that WHAT'S BEST is correct: S desires $_Q$ p if and only if: (i) p is the

most desirable option in Q or (ii) p is the disjunction of the most desirable options in Q . Then we could try to analyze the desirability facts in terms of the desire facts, perhaps as follows:

- $p >_S q$ iff: S wants _{$\{p,q\}$} p . (That is: S prefers _{$\{p,q\}$} p to q .)
- $p =_S q$ iff: S merely wants _{$\{p,q\}$} $p \vee q$.¹⁰¹ (That is: S is indifferent _{$\{p,q\}$} between p and q .)

In fact, supposing S 's (binary) preferences satisfy certain structural constraints (including completeness and transitivity—see [Steele & Stefánsson 2025](#), §§2-3 for an overview), then the desirability ordering with which we associate S could be upgraded to an (expected) utility function. But let's not worry about going that far. The important point is that if WHAT'S BEST were right, then knowing an agent's binary preferences would, in principle, be enough to know the agent's desirability ordering. And this makes the prospect of an analysis of desirability in terms of desire a live option.

But WHAT'S BEST is false. How does the situation change with COGENT DESIRE in view?

Well, nothing stops us from *defining* ' $>_S$ ' and ' $=_S$ ' in this way and pronouncing the corresponding relations as 'being more desirable than' and 'being exactly as desirable as'. But for the reasons given in §6 and §7, I think it's fair to say that the resulting relations are not fit to do the sort of theoretical work desirability orderings are meant to do. Desirability (so understood) will not supervene on any more familiar kinds of value. Someone who has a Buridan preference for Left over Right will automatically have a desirability ordering on which Left is more desirable than Right, despite the fact that Left and Right are qualitative duplicates in all the respects that could possibly matter. And someone who ascetically fails to prefer the free \$1 will automatically have a desirability ordering on which a free \$1 is no more desirable than the status quo, despite the fact that having a free \$1 is better than not having it in all the respects that could possibly matter. And thanks to the possibility of maximally fickle Buridan preferences and moderate asceticism—patterns of desire formation that *appear* rational—the resulting desirability orderings will have no hope of being represented by (expected) utility functions.¹⁰² As I see it, then, any ordering on propositions worthy of the title 'desirability ordering' is one on which, for a relatively typical person S , Left and Right are equally desirable and having a free \$1 is more desirable than not having it.

With this fixed point in mind, we can immediately see that the kind of reduction that works conditional on WHAT'S BEST won't work conditional on COGENT DESIRE. Given COGENT DESIRE, you can prefer _{$\{p,q\}$} p to q even when $p =_S q$; and you can fail to prefer p to q even when $p >_S q$. Thus, *in the general case*, knowing that S prefers _{$\{p,q\}$} p to q is only enough to tell you that $q \not>_S p$; and knowing that S is indifferent _{$\{p,q\}$} between p and q tells you nothing at all about the relative desirability of p and q to S . Again, this is no surprise: the *raison d'être* of COGENT DESIRE is to allow desire to fail to supervene on the desirable.

¹⁰¹ We may need to complicate this condition slightly. An agent who doesn't understand (or hasn't considered) what's at issue between p and q might automatically count as neither preferring _{$\{p,q\}$} p to q nor preferring _{$\{p,q\}$} q to p , but it needn't obviously follow from this that $p =_S q$. There's also the question of whether this reduction papers over the possibility of incomparability (§3.2): one might fail to have a preference _{$\{p,q\}$} between p and q because one regards p and q as incomparably good, rather than equally good. Similar points apply below. I will, however, such matters aside.

¹⁰² Buridan preferences would correspond to violations of the "continuity" requirement, moderate asceticism to violations of the transitivity of indifference.

However, there is a different kind of supervenience thesis in the vicinity: namely, the supervenience of the desires one in a certain sense *could* have on the desirable.¹⁰³ This is because if $p >_S q$, then there are exactly two states of $\text{desire}_{\{p,q\}}$ consistent with the requirements of cogency: S could $\text{desire}_{\{p,q\}}$ p in particular or S could merely $\text{desire}_{\{p,q\}}$ $p \vee q$. And if $p =_S q$, then there are exactly three states of $\text{desire}_{\{p,q\}}$ consistent with the requirements of cogency: S could $\text{desire}_{\{p,q\}}$ p in particular or S could $\text{desire}_{\{p,q\}}$ q in particular or S could merely $\text{desire}_{\{p,q\}}$ $p \vee q$. Similar points apply to more complicated alternative sets. If $p >_S q >_S r$, then there are three cogency-consistent $\text{desire}_{\{p,q,r\}}$ states; if exactly one of the ' $>_S$'s is replaced with a ' $=_S$ ' it becomes four; if both are replaced it becomes seven. With these points in mind, we could try for a modal analysis along the lines of:

- $p >_S q$ iff: S could want $_{\{p,q\}}$ p or could want $_{\{p,q\}}$ $p \vee q$, but could not want $_{\{p,q\}}$ q .
- $p =_S q$ iff: S could want $_{\{p,q\}}$ p or could want $_{\{p,q\}}$ q or could merely want $_{\{p,q\}}$ $p \vee q$.

But there is an immediate issue with this analysis: without restrictions on the interpretation of 'could', we'll get obvious failures in the right-to-left direction. For example: suppose in fact I am indifferent to the prospect of going Left and the prospect of going Right. Here's one way in which it could be that I want to go Left in particular: it could be that I make up my mind that way. Here's another: it could be that the scenario changes, and I learn that the \$20 in Right has been replaced with pocket lint, while the \$20 in Left is upgraded to \$1,000,000. These are both possibilities in which, unlike in actuality, I want to go Left. But intuitively the latter kind of possibility shouldn't suffice to establish that, here in actuality, Left is at least as desirable as Right. This is a reflection of the banal observation that the desirability of a prospect can be a highly contingent matter.

We could get around this sort problem if we restricted the domain of the 'could' to worlds in which S 's desirability ordering is held fixed. But then the analysis is trivial. Is there any reasonable way of replicating this restriction without building the facts about desirability directly into the picture?

I'd be surprised. Notice, for example, that it won't be enough to restrict attention to "nearby" possibilities: (i) the desirability facts are the kinds of facts that could easily be different, on any reasonable conception of what they are; and (ii) supposing the limitations on our powers of voluntary desire formation are moderately counterfactually robust, we'd start getting too many failures in the left-to-right direction.

I see little reason to think that other proposals in this spirit will fare much better. So although I think there are true interpretations of claims like:

¹⁰³ Cf. Gert (2004, pp.493-494) on the relationship between judgments about rationally permissible choices and "valuations":

The central piece of conceptual apparatus... is the distinction between the choices that a person is disposed to make, and the choices that a person is disposed to regard as rationally permissible. The former may plausibly be taken to reveal a person's preferences, but a person's valuations are more plausibly taken to be revealed by the latter. ... For on this picture of human choice and valuation, our valuations of two items are not given by our dispositions to choose, but by our dispositions either to understand or to be puzzled by certain choices: to regard them as rational or irrational, mistaken or not mistaken.

S 's desirability ordering with respect to the propositions in Q at time t is a function of the desires _{Q} S could have had at time t .

I don't see much hope in leveraging them into a reductive analysis of desirability. This is not to say such claims are completely uninformative—here again it is instructive to consider the analogy to the notion of *similarity* in the theory of counterfactuals or *safety* in the theory of knowledge. The point is just that their explanatory value is in giving a sense of the contours of the notion, rather than a reductive analysis of it.

There's also an old worry lurking in the background for any desire-based analysis of desirability, which is that it faces a Euthyphro problem.¹⁰⁴ Intuitively, we desire something because we find it desirable, not the other way around. Now, in the context of COGENT DESIRE, this slogan needs to be complicated a bit: desire is not an *automatic* response to the desirable. But desirability still constrains desire enough to press the basic point: e.g., I am free to choose my preferences in Buridan cases *because* the options are equidesirable, not the other way around. How then could the facts about desire (whether here or in nearby worlds) settle the facts about desirability?

In light of these concerns, I don't think we can reasonably expect to construct the desirability facts out of the desire facts. Might some other kind of analysis fare better? Perhaps instead of trying to reduce the desirability facts to the desire facts, we cut out the middleman and reduce the desirability facts to the kinds of facts some philosophers have taken the facts about *desire* to be reducible to—dispositions to feel pleasure, beliefs about and/or appearances of the good, reasons for acting, and so on.¹⁰⁵ I don't have much to say about this approach, except to emphasize that the problems these sorts of accounts face *qua* theories of the grounds of desire—putative counterexamples, threats of circularity, etc.—do not seem to improve in any meaningful way *qua* theories of the nature of desirability. This is not to say that facts about pleasure, beliefs about the good, reasons for acting, etc., are irrelevant to matters of desirability. Quite to the contrary, I believe the desirability facts depend quite sensitively on the distribution of these sorts of things. But it seems to me no accident that work on desire in the tradition of philosophical semantics so often leaves the notion of desirability as an unanalyzed primitive (see the citations in §8).

I am thus inclined to endorse a kind of non-constructive realism about desirability—i.e., to take the desirability facts to be fundamentally *sui generis*. I imagine there are a number of philosophers who would be uncomfortable adopting this position.¹⁰⁶ Unfortunately, I lack the space to engage with the standard sources of skepticism. But I will conclude this discussion by briefly mentioning one respect in which, conditional on bouletic voluntarism, we face greater pressure to countenance a theoretically robust notion of desirability than we might otherwise.

As we know from (§11), we exhibit some voluntary control over our desires, but not total

¹⁰⁴ *Mutatis mutandis* for those who try to reduce the facts about desirability to the facts about dispositions to act or choose. I'll have more to say about the connections between desire, desirability, and action in §14. For discussion of closely related issues, see, e.g., Smith *et al.* (1989), Kraut (1994), Rabinowicz & Rønnow-Rasmussen (2004), and Schroeder (2004, Ch. 1).

¹⁰⁵ Again see Schroeder (2020, §1) for an overview of some of the options here.

¹⁰⁶ For a small sampling of the issues involved here—though with the caveat that most of these authors are concerned with the question of realism about *utility-grade* desirability orderings—see, e.g.: Allais (1979), Dreier (1996), Broome (1991, 1999), Eriksson & Hájek (2007), Meacham & Weisberg (2011), Buchak (2013), Okasha (2016), Bradley (2017).

voluntary control. The patterns here cry out for explanation. Why is it up to me whether I want a free \$1, but not up to me whether I want a free \$1,000,000? Where exactly is the point where my power to will gives out? Under what conditions would that point lie elsewhere? Without recourse to something akin to desirability, it's not at all clear how we're supposed to progress with questions like this. Indeed, my preferred model of the phenomenon (§11.2) not only requires the existence of desirability facts, but desirability facts with enough structure to allow for meaningful comparisons of *degrees* of desirability. Until we find a better story, I believe we should be prepared to treat desirability orderings, perhaps quite richly structured, as fundamental posits of folk psychology.

13 Rational desire

So far we have concerned ourselves exclusively with questions about the metaphysics of desire. But how should we thinking about *rational* desire, in light of COGENT DESIRE and bouletic voluntarism? This too is an enormous topic. My aim here is to convey some of the natural choice points.

David Hume once said ‘Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger’ (1738, 2.3.3.6). I’m going to operationalize ‘rational desire’ in these terms: for S ’s desire _{Q} that p to be rational is for it not to be contrary to reason—whatever that means—for S to desire _{Q} that p . And I’m going to assume without argument that Hume is right in taking the principles of rational desire to be extremely permissive: there is nothing *irrational* about preferring_{destruction,scratch} the destruction of the world to the scratching of one’s finger.

On the Humean view I am taking for granted, the rational constraints on desire are structural rather than substantive. We can indict someone for having certain *patterns* of desire. But we can’t indict have someone for having any particular desire—at least not in isolation. However, in the context of an alternative-sensitive theory of desire—especially one that allows for failures of the supervenience of desire on the desirable—it’s a somewhat delicate matter saying which patterns of “desire” are or are not contrary reason. But the guiding idea will be this: so long as one’s desirability ordering is rational, any desire _{Q} one could have (given COGENT DESIRE) could in principle be rational. Rational desirability and cogency is necessary and sufficient for *intra- Q* bouletic rationality. However, it is plausible that there are further requirements on *inter- Q* bouletic rationality.

So let’s start with the question of rational *desirability*. What does it take for S ’s desirability ordering $\geq_S$ to not be contrary to reason? Well, at a minimum it seems like it should have a certain degree of coherence. For example, we might require it to obey generalizations like:

- If $p \geq_S q$ and, necessarily, [p iff p^* and q iff q^*], then $p^* \geq_S q^*$.
- If, necessarily, $\neg(p \wedge q)$, and $p \geq_S q$, then $p \geq_S (p \vee q) \geq_S q$.
- If $p >_S q$, and if G_1 is a gamble that gets you p with probability x and q with probability $(1-x)$, while G_2 is a gamble that gets you p with probability $(x-\epsilon)$ and q with probability $(1-(x-\epsilon))$, then $G_1 >_S G_2$.

Indeed, we could cover quite a lot of ground at once by requiring $\succsim_S$ to be a total pre-order (complete and transitive). We could cover even more ground if required $\succsim_S$ to have the structure needed to be representable by an (expected) utility function.¹⁰⁷

Whether these further requirements are plausible depends on difficult issues concerning money-pumps (really: desirability-pumps) and the nature of the values to which desirability is responsive.¹⁰⁸ Is there something contrary to reason about being the kind of person who generally finds it more desirable to have more money than less; finds it desirable to spend \$1 to swap A for B and desirable to spend \$1 to swap B for C; but yet also finds it desirable to spend \$1 to swap C for A? To be such a person is to be someone who could regard each of a series of prospects desirable while also regarding the series as a whole as undesirable (in this case: cycling from A back to itself and ending up \$3 poorer in the process). At first blush this certainly looks contrary to reason! On the other hand, though, it's a bit hard to know what to make of such judgments without a clearer sense on the relationship between desirability and action (more on which soon).

In any event, for the sake of concreteness, I'm going to assume that it's a rational requirement on desirability orderings that they have utility-grade structure. What then are the rational requirements on desire itself, for agents who have rational desirability orderings?¹⁰⁹

Here it will be useful to distinguish *intra*-alternative set rational requirements from *inter*-alternative set rational requirements. We'll start with the former. In both cases we'll be strictly speaking concerned with the rational requirements on *ur*-desire, but for simplicity's sake I'll just pronounce it 'desire'.

So, looking at the options in Q in isolation, what is S rationally permitted to desire $_Q$? The answer I prefer is: any cogent $_{S,Q}$ proposition. The argument for this view is simple: as argued in §§6-7 desire reports that express tie-breaking and/or asceticism are perfectly felicitous. This suggests not just that it's *possible* for one's desires to break desirability ties or to fail to distinguish between options that are unequally desirable, but that it is *not contrary to reason* for one's desires to be this way. As I see it, there's no clear sense in which one is making a *mistake* in having a preference between Left and Right, or in failing to have a preference for a free \$1. This is in much the same way that there is no clear sense in which one is making a mistake in guessing that a fair coin will land heads, or in failing to *strongly* prefer the free \$1, or in failing to *wish* that one had taken it. That's just not how those attitudes work. And although it might be a mistake not to *choose* to have the free \$1, this alone alone does very little to show that it would be a mistake to fail to prefer the free \$1. Indeed, if it is a mistake not to take the free \$1 even

¹⁰⁷ For assumptions about what exactly needs to be assumed about the properties of $\succsim_S$, see Jeffrey (1965), Joyce (1999), Bacon (2022), Nebel (2024), Steele & Stefánsson (2025).

¹⁰⁸ See Gustafsson (2022) and the citations therein for a helpful overview of the issues here.

¹⁰⁹ Assuming it's a rational requirement on desirability orderings that they be representable by (expected) utility functions, there's a real risk that no actual person's overall desirability ordering is rational. This might seem to make the main-text's guiding question a bit remote: shouldn't we be concerned with the question of which states of rational desire are permissible for actual human beings? But one nice feature of the alternative-sensitive framework is that we can make *local* assessments of rational desirability. We can ask, for instance, whether S 's desirability ordering obeys the stated requirements with respect to the propositions in some particular alternative set Q , or the propositions that constitute the union of some collection of alternative sets Q_1 - Q_n (cf. Hoek 2025). This is a much lower bar to clear, and I don't see much downside to the view that an agent whose desirability ordering is irrational with respect to the propositions in Q has irrational desires $_Q$ (or that one whose desirability ordering is irrational with respect to the propositions that are in the union of Q_1 - Q_n has an overall state of desires $_{Q1}$, desires $_{Q2}$, ..., and desires $_{Qn}$ that is irrational).

when fails to have a preference for it, then all that shows is that (rational) action is not settled by matters of (rational) preference. (Much in the same way that (rational) action is not settled by matters of (rational) strong preference, wishing, loving, and so on.) But we'll come back to such matters in the next section.

That covers my preferred view of the rational norms on intra- Q desire: since S 's desire $_Q$ is guaranteed to be cogent $_{S,Q}$, S 's desire $_Q$ is guaranteed to be rational—at least when considered in isolation—so long as S 's desirability ordering is rational.¹¹⁰

Why the caveat about “being considered in isolation”? Because certain patterns of desire across *different* alternative sets look contrary to reason, even when each of those desires taken on its own does not. For example: suppose you're the kind of person who (rationally) prefers a free \$1,000,000 to \$0. Then it seems like there would be something contrary to reason in you *failing* to prefer a free \$1,000,001 to \$0. Likewise for indifference. If you're the kind of person who is (rationally) indifferent between the prospect of a free \$1 and the prospect of \$0, then it seems like there would be something contrary to reason in you have a preference for the prospect of free 75 cents over the prospect of a free 25 cents. But nothing we have said so far delivers these verdicts. The mere fact that (rationally) $x^+ >_S x >_S y >_S y^-$ does nothing to imply that if you prefer $_{\{x,y\}}$ x to y , then you prefer $_{\{x^+,y\}}$ x^+ to y . This is because the cogency $_{S,\{x,y\}}$ of preferring $_{\{x,y\}}$ x to y is compatible with the cogency $_{S,\{x^+,y\}}$ of being indifferent $_{\{x^+,y\}}$ between x^+ and y . (*Mutatis mutandis* in the indifference case.)

To rule out such cases, we need *inter- Q* constraints on rational desire formation. With respect to the first kind of case, we need a principle like:

INCREASED MARGINS

For any $x^+ >_S x$; any Q_1, Q_2 such that $x \in Q_1$, $x^+ \in Q_2$, and $Q_1 \cap Q_2 = Q_1 \setminus \{x\} = Q_2 \setminus \{x^+\}$; and any p_1 such that x entails p_1 and p_2 such that x^+ entails p_2 :

S is rationally required to: ur-desire $_{Q_2}$ p_2 if S ur-desires $_{Q_1}$ p_1 .

What INCREASED MARGINS says, effectively, is this. If you are rational, then if relative to some alternative set Q_1 you want a distinguished option (x), then: relative to Q_2 —the alternative set you get by replacing the distinguished option with an even more distinguished option (x^+)—you also want the more distinguished option (x^+).

And with respect to the second kind of case, we need a principle like::

DECREASED MARGINS

For any $x^+ \geq_S x \geq_S y \geq_S y^-$; any Q_1, Q_2 such that $x^+, y^- \in Q_1$, $x, y \in Q_2$ and $Q_1 \cap Q_2 = Q_1 \setminus \{x^+, y^-\} = Q_2 \setminus \{x, y\}$; and any p_1 such that both or neither x^+ and y^- entail p_1 and p_2 such that both or neither x and y entail p_2 :

S is rationally required to: ur-desire $_{Q_2}$ p_2 if S ur-desires $_{Q_1}$ p_1 .

¹¹⁰ Those who are less sympathetic to the Humean perspective will likely want to deny this. For example: when in a Kantian mood, one might think that it is always contrary to reason to fail to want to what is morally right. We could try to capture some amount of this idea by building it into the desirability ordering—say by insisting that if p is morally right but q is not, then $p >_S q$. But that won't be enough to capture the full force of the idea, since on the view I've defended one is still rationally permitted to be indifferent $_{\{p,q\}}$ between p and q . In a similar vein, one might think that it is contrary to reason to be anything other than indifferent between options associated with moral dilemmas, like in Sophie's choice (Rabinowicz 2008, p. 28). Here too there is no obvious way of constraining the space of rational desirability orderings to force indifference: by the lights of COGENT DESIRE, one is free to form preferences between options that are equally good or incomparably good.

What DECREASED MARGINS says, effectively, is this. If you are rational, then if relative to some alternative set Q_1 you are indifferent between (let's call them) a good option (x^+) and a bad option (y^-)—in the sense that either both or neither of them is compatible with what you want—then: relative to Q_2 —the alternative set you get by replacing the good option with a no worse option (x) and replacing the bad option with a no better option (y)—relative to that new alternative set, you are indifferent between the no worse option (x) and the no better option (y).

Suppose we accept INCREASED MARGINS and DECREASED MARGINS as principles of rational inter- Q desire formation. Is there some more general theory of the rational constraints on inter- Q desire from which they can be derived? Perhaps. But as of yet I have very little sense for what it might look like. The fundamental issue is that the rational requirements on inter- Q desire formation look like they have to be quite weak.

Consider for example:

TRANSITIVE INDIFFERENCE

For any p, q, r and any Q_1 such that $p, q \in Q_1$ and Q_2 such that $q, r \in Q_2$:

S is rationally required to be such that: if S 's ur-desire_{Q_1} is entailed by either both or neither p and q and S 's ur-desire_{Q_2} is entailed by either both or neither q and r , then for any Q_3 such that $p, r \in Q_3$: S 's ur-desire_{Q_3} is entailed by either both or neither p and r .

We know from §7 that this principle is false. It is rationally permissible to be a moderate ascetic (indifferent to small differences) without being a total ascetic (indifferent to any difference). There is nothing contrary to reason in, for any n , being indifferent $_{\{\$n, \$n+1\}}$ between $\$n$ and $\$n+1$, while *not* being indifferent $_{\{\$n, \$n+1,000\}}$ between $\$n$ and $\$n + 1,000$. Supposing S is like this, then S 's overall pattern of desires is a counterexample to TRANSITIVE INDIFFERENCE: for each n , S is indifferent between $\$n$ and $\$n+1$ relative to $\{\$n, \$n+1\}$, but S is not indifferent between $\$0$ and $\$1,000$ relative to $\{\$0, \$1,000\}$.

Indeed, it's not even clear we can impose a requirement of transitivity on one's inter- Q preferences:

TRANSITIVE PREFERENCE

For any p, q, r and any Q_1 such that $p, q \in Q_1$ and Q_2 such that $q, r \in Q_2$:

S is rationally required to be such that: if S 's ur-desire_{Q_1} is entailed by p but not q and S 's ur-desire_{Q_2} is entailed by q but not r , then for any Q_3 such that $p, r \in Q_3$: it is not the case that S 's ur-desire_{Q_3} is entailed by r but not by p .

The transitivity of $>_S$ rules out many of the potential counterexamples, but not all of them. If $p =_S q =_S r$ —i.e., p, q , and r constitute a three-way Buridan case—then, for all that's been said so far, S could rationally prefer $_{\{p,q\}}$ p to q , prefer $_{\{q,r\}}$ q to r , and prefer $_{\{p,r\}}$ r to p . (Notice the shifts in the underlying alternative set.) Would this be contrary to reason? I don't find it so clear. What's wrong with being the kind of person who regards Rock, Paper, Scissors as equidesirable, and who, when prompted, reports the respective two-way preferences for Rock over Scissors, Scissors over Paper, and Paper over Rock? Even supposing you're the kind of agent whose choices are guided by your preferences (rather than, say, the desirabilities), it's not like you're going to foreseeably make a series of bad choices. So long as your desirability ordering is rational, then

any amount of sweetening (say \$1 for playing Rock) will break the cycle. This seems to accord with reason just fine.

Without TRANSITIVE INDIFFERENCE and TRANSITIVE PREFERENCE, it is difficult to see there being many simple, general coherence constraints on one's overall desire state. This is because without these principles, there is no hope of reading off the facts about what *S* is permitted desire relative to three-or-more-cell alternative sets from *S*'s simple, binary preferences (i.e., *S*'s ur-desires relative to two-cell alternative sets). *S* can be rationally indifferent between (say) \$499 and \$500 relative to {\$499, \$500}, while rationally having a preference between them relative to {\$1, \$2, ..., \$1,000}. And *S* can rationally desire Rock in particular relative to {Rock, Scissors}, while rationally desiring Scissors in particular relative to {Rock, Paper, Scissors}. This is not a situation that looks amenable to the discovery of strong, simple principles.

This is not to say things are *entirely* hopeless. We know from INCREASED MARGINS that if *S* prefers Rock relative to {Rock, Scissors} then *S* prefers Rock relative to {Rock+\$1, Scissors}. And we know from DECREASED MARGINS that if *S* is indifferent between \$499 and \$500 relative to {\$499, \$500}, then *S* is indifferent between \$499.5 and \$500 relative to {\$499.5, \$500}. But I leave the task of finding stronger inter-*Q* constraints on rational desire (supposing there are any) to further investigation.

14 Desire in action

At last we come to the topic of the role of desire in action. My discussion of this topic will make two general assumptions.

First, I will assume that there is an admissible interpretation of 'want' on which it expresses something we theorists might pronounce as '*all-things-considered* desire'.¹¹¹ Presumably this is an interpretation of 'want' on which the operative desirability ordering is *all-things-considered* desirability.¹¹² The person who speaks truly in saying both 'I want to pay my taxes' and 'I want not to pay my taxes' expresses their all-things-considered desires on at most one of these occasions; supposing they in fact pay their taxes, presumably it's on the former occasion. Some assumption of this form seems to me needed if there is to be any hope that slogans like 'We act on the basis of our desires' or 'Desires dispose us to act' express non-trivial truths. So from here on out all talk of desire and desirability is intended to be understood in the "all-things-considered" way.

Second, for convenience I will assume that the objects of actions are propositions (you act to make *p* true). And I will assume that action is a relation between agents and propositions *relative to set of propositions*—intuitively, the options in the relevant choice situation. In other words, then, I will assume that, like desire, action is alternative-sensitive. '*S* chooses_{*Q*} *p*' will abbreviate '*S*'s choice of action when faced with the options in *Q* is *p*'. Not much is going to turn on the claim that action itself is alternative-sensitive, but it will make it a lot easier to state and assess various principles connecting desire (whose alternative-sensitivity matters a great deal) and action.

¹¹¹ For relevant discussion, see, e.g., Davidson (1969), Watson (1975), Davis (1984), Levinson (2003), Condoravdi & Lauer (2016), Phillips-Brown (forthcoming).

¹¹² Strictly speaking it's not *an* interpretation of 'want', but rather a family of such interpretations (one per alternative set).

14.1 Desire and action, descriptively

I'll start with the descriptive connections between desire and action.

Suppose when faced with a choice between p and q , S chooses p . Does it follow from the fact that S chooses_{ p,q } p that S wants_{ p,q } p ?

According to one popular view, the answer is yes:¹¹³

DO WHAT YOU WANT

If S chooses _{Q} p , then S wants _{Q} p .

There is an intuitive case to be made for DO WHAT YOU WANT. For one, it is naturally suggested by the influential idea, traced back to Davidson (1963), that action is behavior is caused by desire (in the right sort of way). It also draws support from judgments of ordinary language. It is standard to move seamlessly from talk of what someone intentionally did to talk of what that person wanted do: witness the standard (albeit sometimes impolite) practice of asking 'Why did you want to do *that*?' after seeing some do something.

However, some have worried that the possibility of choice in Buridan cases undermines DO WHAT YOU WANT (Bratman 1987, 1999, Chislenko 2016, Coren forthcoming). Put roughly, the main argument is this: one can choose_{ p,q } p rather than q even when p and q are equidesirable; one cannot want_{ p,q } p rather than q when p and q are equidesirable (one can only want_{ p,q } $p \vee q$ in such circumstances); so, contrary to DO WHAT YOU WANT, one can choose_{ p,q } p rather than q without wanting_{ p,q } p rather than q .

Given COGENT DESIRE, it is clear that this argument is unsound: p and q being equidesirable is no barrier to one's wanting_{ p,q } p in particular. But I don't think this shows that DO WHAT YOU WANT is safe from the threat of Buridan cases. For although it is *possible* to form discriminating desires in Buridan cases, it is by no means mandatory. And I see no reason to think that someone who fails to form a discriminating desire in a Buridan case will thereby be rendered incapable of action. I can go Left not because I want the contents of Left in particular, but merely because I want the contents of at least one of the two boxes, and going Left is a way of doing that.¹¹⁴ Likewise, supposing I do in fact go Left, a question of the form 'Why do you want to go left rather than right?' seems ripe for being challenged as having a false presupposition. Indeed, a reply along the lines of the following strikes me as entirely appropriate:

'There is no sense whatsoever in which I wanted to go Left rather than Right; all I wanted was to go is some way or other. But the only way to go some way or other was either to go Left or to go Right, and I happened to decide to go Left'.

Similar lessons can be drawn from cases of asceticism. The person who is indifferent to the prospect of a free \$1 is not thereby rendered incapable of acting when they are offered the choice between taking or leaving the free \$1. And even supposing they do take the free \$1, it doesn't follow that they changed their mind about whether they wanted the \$1, or that they must have

¹¹³ For related discussion, both sympathetic and unsympathetic, see e.g., Davidson (1969, 1978), Nagel (1970), Lewis (1974), Davis (1984), Stalnaker (1984), Smith (1987), Clarke (1994), Gert (2005), and Condoravdi & Lauer (2016).

¹¹⁴ By analogy: if there's enough money in both Left and Right, it may be that I *strongly* desire that I get one of Left or Right. But it doesn't follow that I thereby strongly desire to go Left *and* strongly desire to go Right, or that if I do end up going Left it thereby follows that I strongly desired to go Left. Facts about what I strongly desire have no essential connection to my decision to go Left in particular. The same it seems, is true of vanilla desire.

been mistaken or lying when they said they didn't have a preference. An explanation along the following lines once again seems to me entirely appropriate:

'I didn't take the \$1 because I wanted to take it; I took it because I had to either take or decline it, and I didn't care which it was, so I went with taking it.'

Indeed, as far as I can tell, the person who (sincerely) reports not having a preference for the free \$1 might well decide to *decline* it. They are indifferent between the prospect of \$1 and \$0, after all. So why should we expect them to choose the \$1 as a way of resolving their forced choice? (This is not yet to say anything about the *rationality* of such a choice, however.)

The lesson of these points is that if COGENT DESIRE is correct, then if there are any robust principles connecting (all things considered) desire to action, it seems they will have to be of the following form (cf. [Slote 1989](#), [Greenspan 2009](#)):

DO WHAT'S COMPATIBLE

If S chooses _{Q} p , then p is compatible with what S wants _{Q} .

Suppose that's right: given COGENT DESIRE, DO WHAT'S COMPATIBLE is the strongest principle connecting action to (all-things-considered) desire. Does this imply a kind of skepticism about our ability to infer a person's desires from their behavior? Not by itself. It's one thing to acknowledge the metaphysical possibility that someone chooses to accept a free \$1,000,000 not because they want the free \$1,000,000, but because they're indifferent to it and have to either accept it or decline it. It's another to think this sort of thing happens here in actuality, or even anywhere in our modal neighborhood. Speaking for myself, I know that when I'm offered a free \$1,000,000, I'm just going to take it, and I'm going to take it because I prefer to take it. I know that this fact about me is fairly counterfactually robust. And I have every reason to believe that I'm not atypical in this respect. Thus, so long as I'm in the good case, I'm going to be in a pretty good position to know that you are going to want a free \$1,000,000, and that when it is offered to you you will take it because they prefer to. Of course, as the options get closer together—as the \$1,000,000 shrinks to \$1,000 or \$100 or \$1—I'm going to be less well-positioned to infer your desires from your behavior. But as the arguments of §§6–7 suggest, this is a feature of COGENT DESIRE, not a bug.

14.2 Desire and action, rationally

What about the connections between rational desire and rational action? Here I see two general view one could take, the discussion of which will conclude the paper.

The first is what I will call the *desire first* view. On this view, one's actions are rational so long as: (i) one acts in accordance with one's desires; and (ii) one's desires are rational. To satisfy (i), one must abide by DO WHAT'S COMPATIBLE: that is, it must be that one chooses _{Q} p only if p is compatible with one's desires _{Q} . To satisfy (ii), one must abide by the rational norms on desire. Now, §13 did not establish anything resembling a definitive account of what those norms are. But for now let's work with a simple picture on which the only norms on rational desire are derivative on the norms on rational desirability, which (the norms on rational desirability)

amount to the requirement that one's desirability ordering be representable by an (expected) utility function.¹¹⁵

The main argument for the desire first view, at least as I see it, boils down to the following kind of thought: if everything I'm doing is consistent with what I want to do, and if everything I want is rational for me to want, then what mistake could I be making in doing what I do? If I really am indifferent between a free \$1 and the status quo, in what sense am I exhibiting a failure in rationality in failing to pursue it? Yes, I'll get less of what I find desirable than I would have had I chosen to pursue it, but I don't *care* about such small differences. Tis not contrary to reason to be indifferent to small differences in desirability.

The main argument against the desire first view is that it is mad. *Total asceticism*—that is, complete indifference_Q relative to every alternative set Q —is a rationally permissible overall state of desire. So given DO WHAT'S COMPATIBLE, it follows that quite literally any pattern of behavior on the part of S could, in principle, be rational—even if S 's desirability ordering is exactly like yours or mine. S just has to be sufficiently ascetic. Now, as we know from §11, no actually existing person is capable of such impressive feats of asceticism. But even focusing on agents (like you and me) who are incapable of total asceticism (say because $\Delta_S \approx 0$), the situation looks grim. For so long as S is capable of *some* amount of asceticism, the desire first view predicts that S can rationally and knowably exhibit a certain kind of *akrasia*: S can intentionally p rather than q , even while knowing that q is better than p in all possible respects. Unsurprisingly, it also predicts that S can willingly be money-pumped and be rational for it. Finally, there is the fact that, given bouletic voluntarism, the desire first view implies that S can make it easier or harder for themselves to act in accordance with reason just by deciding to desire differently. If S chooses to become pickier, it becomes harder for them to act rationally; if S chooses to become less picky, it becomes easier to act rationally. When in doubt, be indifferent, and then do the thing you would have done had you not been indifferent!

Someone who is attracted to the desire first view could avoid a number of these problems by imposing more stringent requirements (presumably both intra- Q and inter- Q) on rational desire formation. But to get around these problems they'd have to be quite stringent indeed. And there is a worry that any such requirements will fly in the face of the way we ordinarily think and talk about desire, suggesting that they are rational requirements not of (ordinary) desire, but of some desire-adjacent attitude.

A better option, then, might just be to adopt the second view I want to consider: the desirability first (desire last) view. On this view, one's actions are rational so long as: (i) one acts in accordance with one's desirability ordering; and (ii) one's desirability ordering is rational. We know what it takes to satisfy (ii). What about (i)? Here there are a few options, but for concreteness I'll focus on:

DO WHAT'S BEST

If S chooses_Q p , then for any $q \in Q$, $p \not\prec_S q$.

On this view, rational agents act like (expected) utility maximizers. Any two agents who are exactly alike in their desirability orderings are exactly alike in which actions they are rationally

¹¹⁵ In other words: for now we're assuming there are no inter- Q requirements on rational desire.

permitted to take. Supposing they are in fact rational, they will also be exactly alike in which actions they *will* take, except in Buridan cases (and perhaps with respect to which first-personal desire reports they will assert).

The main argument for the desirability first (desire last) view is that it avoids the unsavory consequences of the desire first view. I consider this a very strong argument indeed. Of course, the desirability first view is not without its downsides—see any of the vast number of criticisms of standard decision-theory as a normative theory of action. But we could, in principle, mitigate at least some of these problems by replacing DO WHAT’S BEST with a weaker principle (say, DO WHAT’S GOOD ENOUGH, in the manner of [Slote 1989](#)), or by relaxing the requirements on rational desirability orderings. Either way these kinds of adjustments won’t affect the basic situation, which is this: (i) the fundamental norm on desire is cogency, which is highly permissive; (ii) the facts about one’s desires can be determined by arbitrary voluntary selection; and (iii) nothing like (i) or (ii) seems to be true of desirability. Given all that, anything within the spirit of the desirability first view is bound to look much closer to a reasonable theory of rational action than anything within the spirit of the desire first view.

I am thus inclined to draw the following conditional conclusion: if the core theses of this paper are correct—if bouletic voluntarism and COGENT DESIRE are both true—then desire is not what matters to rational decision making. I leave it to the reader to decide whether they prefer accepting the conclusion or rejecting the antecedent.

References

- Al-Ghazali. 1095. *Tahafut Al-Falasifah*. Lahore: Pakistan Philosophical Congress, 1963.
- Aldred, Jonathan. 2007. Intransitivity and Vague Preferences. *The Journal of Ethics*, **11**(4), 377–403.
- Allais, Maurice. 1979. The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School. Pages 27–145 of: Allais, Maurice, & Hagen, Ole (eds), *Expected Utility Hypotheses and the Allais Paradox*. Dordrecht: D. Reidel. Translation of “Fondements d’une Théorie Positive des Choix Comportant un Risque et Critique des Postulats et Axiomes de l’École Américaine,” originally published 1952.
- Alston, William P. 1988. The Deontological Conception of Epistemic Justification. *Philosophical Perspectives*, **2**, 257–299.
- Anand, Paul. 1993. *Foundations of Rational Choice Under Risk*. Oxford University Press.
- Andersson, Henrik, & Hsieh, Nien-hê. 2025. Incommensurable Values. In: Zalta, Edward N., & Nodelman, Uri (eds), *The Stanford Encyclopedia of Philosophy*, Fall 2025 edn. Metaphysics Research Lab, Stanford University.
- Armstrong, W. E. 1950. A Note on the Theory of Consumer’s Behaviour. *Oxford Economic Papers*, **2**(1), 119–122.
- Audi, Robert. 2001. *The Architecture of Reason: The Structure and Substance of Rationality*. New York: Oxford University Press.
- Bacon, Andrew. 2022. Actual Value in Decision Theory. *Analysis*, **82**(4), 617–629.
- Baker, Derek. 2010. Ambivalent Desires and the Problem with Reduction. *Philosophical Studies*, **150**(1), 37–47.
- Barrett, Jacob. 2019. Interpersonal Comparisons with Preferences and Desires. *Politics, Philosophy and Economics*, **18**(3), 219–241.
- Baumann, Peter. 2022. Rational Intransitive Preferences. *Politics, Philosophy and Economics*, **21**(1), 3–28.
- Bayle, Pierre. 1697. *Dictionnaire historique et critique, par Monsieur Bayle*. Rotterdam: R. Leers.
- Beck, Lukas. forthcoming. Desires in Economics. In: Gregory, Alex (ed), *The Routledge Handbook on the Philosophy of Desire*. Routledge.
- Bermúdez, José Luis. 2009. *Decision Theory and Rationality*. Oxford, GB: Oxford University Press.
- Blumberg, Kyle. 2022a. On Preferring. *Linguistics and Philosophy*, **45**(6), 1315–1344.
- Blumberg, Kyle. 2022b. A Problem for the Ideal Worlds Account of Desire. *Analysis*, **82**(1), 7–15.
- Blumberg, Kyle, & Hawthorne, John. 2021. Wanting What’s Not Best. *Philosophical Studies*, **179**(4), 1275–1296.
- Blumberg, Kyle, & Hawthorne, John. 2022. Desire. *Philosophers’ Imprint*, **22**(n/a).
- Blumberg, Kyle, & Holguín, Ben. 2019. Embedded Attitudes. *Journal of Semantics*, **36**(3), 377–406.
- Boespflug, Mark, & Jackson, Elizabeth. 2024. Doxastic Voluntarism. In: Zalta, Edward N., & Nodelman, Uri (eds), *The Stanford Encyclopedia of Philosophy*, Winter 2024 edn. Metaphysics Research Lab, Stanford University.
- Booth, Richard Jefferson. 2023. Underspecifying Desires. *Linguistics and Philosophy*, **5**, 1–30.
- Bradley, Richard. 1999. Conditional Desirability. *Theory and Decision*, **47**(1), 23–55.
- Bradley, Richard. 2017. *Decision Theory with a Human Face*. Cambridge University Press.
- Bratman, Michael. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Bratman, Michael. 1999. Davidson’s Theory of Intention. Pages 209–224 of: *Faces of Intention: Selected Essays on Intention and Agency*. New York: Cambridge University Press.
- Braun, David. 2015. Desiring, Desires, and Desire Ascriptions. *Philosophical Studies*, **172**(1), 141–162.
- Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Wiley-Blackwell.
- Broome, John. 1999. *Ethics Out of Economics*. New York: Cambridge University Press.
- Buchak, Lara. 2013. *Risk and Rationality*. Oxford, GB: Oxford University Press.
- Carter, Sam, & Hawthorne, John. forthcoming. Normality. *Journal of Philosophy*.
- Chang, Ruth. 2004. Can Desires Provide Reasons for Action. Pages 56–90 of: Wallace, R. Jay (ed), *Reason and value: themes from the moral philosophy of Joseph Raz*. Oxford University

- Press.
- Chang, Ruth. 2009. Voluntarist Reasons and the Sources of Normativity. *Pages 243–71 of: Sobel, David, & Wall, Steven (eds), Reasons for Action.* Cambridge University Press.
- Chang, Ruth. 2013a. Commitments, Reasons, and the Will. *Oxford Studies in Metaethics*, **8**.
- Chang, Ruth. 2013b. Grounding Practical Normativity: Going Hybrid. *Philosophical Studies*, **164**(1), 163–187.
- Chang, Ruth. 2017. Hard Choices. *Journal of the American Philosophical Association*, **3**(1), 1–21.
- Chang, Ruth. 2020. Do We Have Normative Powers? *Aristotelian Society Supplementary Volume*, **94**(1), 275–300.
- Chislenko, Eugene. 2016. A Solution for Buridan's Ass. *Ethics*, **126**(2), 283–310.
- Chomsky, Noam. 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought.* Cambridge University Press.
- Clarke, Randolph. 1994. Doing What One Wants Less: A Reappraisal of the Law of Desire. *Pacific Philosophical Quarterly*, **75**(1), 1–11.
- Condoravdi, Cleo, & Lauer, Sven. 2016. Anakastic conditionals are just conditionals. *Semantics and Pragmatics*, **9**(8), 1–69.
- Coren, Daniel. forthcoming. Equal Desires and Self-Control. *Inquiry: An Interdisciplinary Journal of Philosophy*.
- Crnić, Luca. 2011. *Getting even.* Ph.D. thesis, MIT.
- Danielsson, Sven. 1998. Numerical Representations of Value-Orderings: Some Basic Problems. *Pages 19–114 of: Fehige, Christoph, & Wessels, Ulla (eds), Preferences.* Walter de Gruyter, Berlin.
- Danto, Arthur C. 1965. Basic Actions. *American Philosophical Quarterly*, **2**(2), 141–148.
- Davidson, Donald. 1963. Actions, Reasons, and Causes. *Journal of Philosophy*, **60**(23), 685.
- Davidson, Donald. 1969. How is Weakness of the Will Possible? *In: Feinberg, Joel (ed), Moral concepts.* Oxford University Press.
- Davidson, Donald. 1970. Mental Events. Clarendon Press.
- Davidson, Donald. 1978. Intending. *Philosophy of History and Action*, **11**, 41–60.
- Davidson, Donald, McKinsey, John, & Suppes, Patrick. 1955. Outlines of a Formal Theory of Value, I. *Philosophy of Science*, **22**(2), 140–160.
- Davis, Wayne A. 1984. The Two Senses of Desire. *Philosophical Studies*, **45**(2), 181–195.
- Dietrich, Franz, & List, Christian. 2016. Mentalism Versus Behaviourism in Economics: A Philosophy-of-Science Perspective. *Economics and Philosophy*, **32**(2), 249–281.
- Dorr, Cian, Nebel, Jacob M., & Zuehl, Jake. 2023. The Case for Comparability. *Noûs*, **57**(2), 414–453.
- Dorst, Kevin, & Mandelkern, Matthew. 2023. Good Guesses. *Philosophy and Phenomenological Research*, **105**(3), 581–618.
- Dreier, James. 1996. Rational Preference: Decision Theory as a Theory of Practical Rationality. *Theory and Decision*, **40**(3), 249–276.
- Elson, Luke. 2021. What Does Incommensurability Tell Us About Agency? (1st edition). *Pages 181–198 of: Andersson, Henrik, & Herlitz, Anders (eds), Value Incommensurability: Ethics, Risk. And Decision-Making.* Routledge.
- Eriksson, Lina, & Hájek, Alan. 2007. What Are Degrees of Belief. *Studia Logica*, **86**(2), 185–215.
- Fanciullo, James. 2022. On Sense and Preference. *Journal of Moral Philosophy*, **19**(3), 280–302.
- Fanciullo, James. forthcominga. Decision Theory and Desire. *In: Gregory, Alex (ed), The Routledge Handbook on the Philosophy of Desire.* Routledge.
- Fanciullo, James. forthcomingb. Preference as Desire. *Journal of Philosophy*.
- Fischer, John Martin. 1994. *The Metaphysics of Free Will: An Essay on Control.* Cambridge, Mass.: Wiley-Blackwell.
- Fishburn, P.C. 1973. *The Theory of Social Choice.* Princeton University Press, Princeton.
- Frankfurt, Harry G. 1969. Alternate Possibilities and Moral Responsibility. *Journal of Philosophy*, **66**(23), 829–839.
- Frankfurt, Harry G. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, **68**(1), 5–20.
- Gendin, Sidney. 1996. Why Preference is Not Transitive. *Philosophical Quarterly*, **46**(185), 482–488.
- Gert, Joshua. 2004. Value and Parity. *Ethics*, **114**(3), 492–510.

- Gert, Joshua. 2005. Breaking the Law of Desire. *Erkenntnis*, **62**(3), 295–319.
- Ginet, Carl. 1966. Might We Have No Choice. *Pages 87–104 of: Lehrer, Keith (ed), Freedom and Determinism. Contributors: Roderick M. Chisholm And Others.* Random House.
- Goodman, Jeremy, & Salow, Bernhard. 2023. Epistemology Normalized. *Philosophical Review*, **132**(1), 89–145.
- Grano, Thomas, & Phillips-Brown, Milo. 2022. (Counter)Factual Want Ascriptions and Conditional Belief. *Journal of Philosophy*, **119**(12), 641–672.
- Greenspan, Patricia. 1980. A Case of Mixed Feelings: Ambivalence and the Logic of Emotion. *Pages 223–250 of: Rorty, Amélie Oksenberg (ed), Explaining Emotions.* University of California Press.
- Greenspan, Patricia. 2009. Resting Content: Sensible Satisficing? *American Philosophical Quarterly*, **46**(4), 305–317.
- Guala, Francesco. 2019. Preferences: Neither Behavioural nor Mental. *Economics and Philosophy*, **35**(3), 383–401.
- Gustafsson, Johan E. 2016. Money Pumps, Incompleteness, and Indeterminacy. *Philosophy and Phenomenological Research*, **92**(1), 60–72.
- Gustafsson, Johan E. 2022. *Money-Pump Arguments.* Cambridge: Cambridge University Press.
- Handfield, Toby. 2014. Rational Choice and the Transitivity of Betterness. *Philosophy and Phenomenological Research*, **89**(3), 584–604.
- Hansson, Sven Ove. 1993. Money-Pumps, Self-Torturers and the Demons of Real Life. *Australasian Journal of Philosophy*, **71**(4), 476–485.
- Harman, Gilbert. 1976. Practical Reasoning. *The Review of Metaphysics*, **29**(3), 431–463.
- Harman, Gilbert. 1998. The Toxin Puzzle. *Pages 84–89 of: Coleman, Jules L., & Morris, Christopher W. (eds), Rational Commitment and Social Justice: Essays for Gregory Kavka.* Cambridge University Press.
- Hausman, Daniel M. 2011. *Preference, Value, Choice, and Welfare.* Cambridge University Press.
- Heim, Irene. 1992. Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics*, **9**(3), 183–221.
- Hoek, Daniel. 2025. Minimal Rationality and the Web of Questions. *In: van Elswyk, Peter, Kindermann, Dirk, Kirk-Giannini, Cameron Domenico, & Egan, Andy (eds), Unstructured Content.* Oxford University Press.
- Holguín, Ben. 2022. Thinking, Guessing, and Believing. *Philosophers' Imprint*, **22**, 1–34.
- Holguín, Ben, & Lederman, Harvey. forthcoming. Trying Without Fail. *Philosophical Studies*, 1–28.
- Holton, Richard. 2009. *Willing, Wanting, Waiting.* Oxford University Press UK.
- Horowitz, Sophie. 2019. Accuracy and Educated Guesses. *In: Gendler, Tamar Szabó, & Hawthorne, John (eds), Oxford Studies in Epistemology 6.* Oxford Scholarship Online.
- Hume, David. 1738. *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects.* Oxford, GB: Oxford University Press.
- Icard, Thomas. 2016. Subjective Probability as Sampling Propensity. *Review of Philosophy and Psychology*, **7**(4), 863–903.
- Icard, Thomas. 2021. Why Be Random? *Mind*, **130**(517), 111–139.
- James, William. 1890. *The Principles of Psychology.* London, England: Dover Publications.
- Jeffrey, Richard C. 1965. *The Logic of Decision.* New York, NY, USA: University of Chicago Press.
- Jerzak, Ethan. 2019. Two Ways to Want? *Journal of Philosophy*, **116**(2), 65–98.
- Joyce, James M. 1999. *The Foundations of Causal Decision Theory.* Cambridge University Press.
- Kavka, Gregory S. 1983. The Toxin Puzzle. *Analysis*, **43**(1), 33–36.
- Kraut, Richard. 1994. Desire and the Human Good. *Proceedings and Addresses of the American Philosophical Association*, **68**(2), 315.
- Lassiter, Daniel. 2011. *Measurement and Modality: the Scalar Basis of Modal Semantics.* Ph.D. thesis, New York University.
- Lehrer, Keith, & Wagner, Carl. 1985. Intransitive Indifference: The Semi-Order Problem. *Synthese*, **65**(2), 249–256.
- Leibniz, Gottfried Wilhelm. 1952. *Theodicy.* Yale University Press. trans. E.M. Hubbard.
- Levi, Isaac. 1986. *Hard Choices: Decision Making Under Unresolved Conflict.* New York: Cambridge University Press.
- Levinson, Dmitry. 2003. Probabilistic model-theoretic semantics for ‘want’. *Semantics and Lin-*

- guistics Theory (SALT), **13**, 222–239.
- Levy, Yair. forthcoming. Attention and Voluntariness in the Wandering Mind. *Journal of Philosophy*.
- Lewis, David. 1974. Radical Interpretation. *Synthese*, **27**(July-August), 331–344.
- Lewis, David K. 1973. *Counterfactuals*. Blackwell.
- Luce, R. Duncan. 1956. Semiorders and a Theory of Utility Discrimination. *Econometrica*, **24**(2), 178–191.
- Luce, Robert Duncan, & Raiffa, Howard. 1957. *Games and Decisions: Introduction and Critical Survey*. New York: Wiley.
- MacKay, A. F. 1980. *Arrow's Theorem: The Paradox of Social Choice*. Yale University Press, New Haven.
- Marino, Patricia. 2009. On Essentially Conflicting Desires. *Philosophical Quarterly*, **59**(235), 274–291.
- Meacham, Christopher, & Weisberg, Jonathan. 2011. Representation Theorems and the Foundations of Decision Theory. *Australasian Journal of Philosophy*, **89**(4), 641–663.
- Mele, Alfred R. 1991. Motivational Ties. *Journal of Philosophical Research*, **16**, 431–442.
- Mele, Alfred R. 2017. Direct Control. *Philosophical Studies*, **174**(2), 275–290.
- Montague, Michelle. 2007. Against Propositionalism. *Noûs*, **41**(3), 503–518.
- Montaigne, Michel Eyquem. 1958. How our mind hinders itself. In: *The Complete Essays of Montaigne*. Stanford University Press. trans. D.M. Frame.
- Nagel, Thomas. 1970. *The Possibility of Altruism*. Oxford: Clarendon P.
- Nebel, Jacob. 2018. The Good, the Bad, and the Transitivity of Better Than. *Noûs*, **52**(4), 874–899.
- Nebel, Jacob. 2024. Ethics Without Numbers. *Philosophy and Phenomenological Research*, **108**(2), 289–319.
- Nichols, Shaun. 2015. *Bound: Essays on Free Will and Responsibility*. Oxford, GB: Oxford University Press.
- Okasha, Samir. 2016. On the Interpretation of Decision Theory. *Economics and Philosophy*, **32**(3), 409–433.
- O'Shaughnessy, Brian. 1973. Trying (As the Mental “Pineal Gland”). *Journal of Philosophy*, **70**(13), 365–386.
- O'Shaughnessy, Brian. 1980. *The Will: A Dual Aspect Theory*. Cambridge University Press.
- Pettit, Philip. 1991. Decision Theory and Folk Psychology. Pages 147–175 of: Bacharach, Michael, & Hurley, Susan (eds), *Essays in the Foundations of Decision Theory*. Blackwell.
- Pettit, Philip. 2002. *Rules, Reasons, and Norms: Selected Essays*. Clarendon Press.
- Phillips-Brown, Milo. 2021. What Does Decision Theory Have to Do with Wanting? *Mind*, **130**(518), 413–437.
- Phillips-Brown, Milo. 2025. Desiderative Lockeanism. *Australasian Journal of Philosophy*, **103**(3), 740–755.
- Phillips-Brown, Milo. forthcoming. Some-Things-Considered Deside. *Semantics and Pragmatics*.
- Pickard, Hanna. 2013. Psychopathology and the Ability to Do Otherwise. *Philosophy and Phenomenological Research*, **90**(1), 135–163.
- Pietroski, Paul M. 2000. *Causing Actions*. Oxford University Press.
- Pollock, John L. 2006. *Thinking About Acting: Logical Foundations for Rational Decision Making*. US: Oxford University Press.
- Quillien, Tadeo, Bramley, Neil R., & Lucas, Christopher G. 2023. *Guesses as Compressed Probability Distributions*.
- Quinn, Warren S. 1990. The Puzzle of the Self-Torturer. *Philosophical Studies*, **59**(1), 79–90.
- Rabinowicz, Wlodek. 2008. Value Relations. *Theoria*, **74**(1), 18–49.
- Rabinowicz, Wlodek, & Rønnow-Rasmussen, Toni. 2004. The Strike of the Demon: On Fitting Pro-Attitudes and Value. *Ethics*, **114**(3), 391–423.
- Rachels, Stuart. 1998. Counterexamples to the Transitivity of Better Than. *Australasian Journal of Philosophy*, **76**(1), 71–83.
- Rachels, Stuart. 2001. A Set of Solutions to Parfit's Problems. *Noûs*, **35**(2), 214–238.
- Raz, Joseph. 1998. Incommensurability and Agency. Pages 110–28 of: Raz, Joseph (ed), *Engaging Reason: On the Theory of Value and Action*. Oxford University Press.
- Raz, Joseph. 1999. *Engaging Reason: On the Theory of Value and Action*. Oxford, UK: Oxford

- University Press.
- Rescher, Nicholas. 1960. Choice Without Preference. A Study of the History and of the Logic of the Problem of “Buridan’s Ass”. *Kant-Studien*, **51**(1-4), 142–175.
- Rubinstein, Aynat. 2017. Straddling the line between attitude verbs and necessity modals. In: *Modality Across Syntactic Categories*. Oxford University Press.
- Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge: Belknap Press of Harvard University Press.
- Schick, Frederic. 1969. Arrow’s Proof and the Logic of Preference. *Philosophy of Science*, **36**(2), 127–144.
- Schopenhauer, Arthur. 1960. *On the Freedom of the Will*. New York, NY, USA: Blackwell.
- Schroeder, Timothy. 2004. *Three Faces of Desire*. New York, US: Oxford University Press.
- Schroeder, Timothy. 2010. Desire and Pleasure in John Pollock’s Thinking About Acting. *Philosophical Studies*, **148**(3), 447–454.
- Schroeder, Timothy. 2020. Desire. In: Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, Summer 2020 edn. Metaphysics Research Lab, Stanford University.
- Schulz, Armin W. 2015. Preferences Vs. Desires: Debating the Fundamental Structure of Conative States. *Economics and Philosophy*, **31**(2), 239–257.
- Schumm, George F. 1987. Transitivity, Preference and Indifference. *Philosophical Studies*, **52**(3), 435–437.
- Sen, Amartya. 1969. Quasi-Transitivity, Rational Choice and Collective Decisions. *The Review of Economic Studies*, **36**(3), 381–393.
- Shemmer, Yonatan. 2004. Desiring at Will and Humeanism in Practical Reason. *Philosophical Studies*, **119**(3), 265–294.
- Shemmer, Yonatan. 2005. Instrumentalism and Desiring at Will. *Canadian Journal of Philosophy*, **35**(2), 269–288.
- Shepherd, Joshua. 2018. Intending, Believing, and Supposing at Will. *Ratio*, **31**(3), 321–330.
- Skipper, Mattias. 2023. Good Guesses as Accuracy-Specificity Tradeoffs. *Philosophical Studies*, **180**(7), 2025–2050.
- Slote, Michael. 1989. *Beyond Optimizing: A Study of Rational Choice*. Cambridge, Mass.: Harvard University Press.
- Smith, Michael. 1987. The Humean Theory of Motivation. *Mind*, **96**(381), 36–61.
- Smith, Michael, Lewis, David, & Johnston, Mark. 1989. Dispositional Theories of Value. *Aristotelian Society Supplementary Volume*, **63**(1), 89–174.
- Sripada, Chandra. 2012. How is Willpower Possible? The Puzzle of Synchronic Self-Control and the Divided Mind. *Noûs*, **48**(1), 41–74.
- Sripada, Chandra Sekhar. 2010. Philosophical Questions About the Nature of Willpower. *Philosophy Compass*, **5**(9), 793–805.
- Stalnaker, Robert. 1968. A Theory of Conditionals. *American Philosophical Quarterly*, 98–112.
- Stalnaker, Robert C. 1984. *Inquiry*. MIT Press.
- Steele, Katie, & Stefánsson, H. Orri. 2025. Decision Theory. In: Zalta, Edward N., & Nodelman, Uri (eds), *The Stanford Encyclopedia of Philosophy*, Fall 2025 edn. Metaphysics Research Lab, Stanford University.
- Stocker, Michael. 1979. Desiring the Bad: An Essay in Moral Psychology. *Journal of Philosophy*, **76**(12), 738–753.
- Strawson, Galen. 2003. Mental Ballistics or the Involuntariness of Spontaneity. *Proceedings of the Aristotelian Society*, **103**(3), 227–256.
- Tanyi, Attila. 2011. Desires as Additional Reasons? The Case of Tie-Breaking. *Philosophical Studies*, **152**(2), 209–227.
- Temkin, Larry S. 1987. Intransitivity and the Mere Addition Paradox. *Philosophy and Public Affairs*, **16**(2), 138–187.
- Temkin, Larry S. 1996. A Continuum Argument for Intransitivity. *Philosophy and Public Affairs*, **25**(3), 175–210.
- Thoma, Johanna. 2020. Folk Psychology and the Interpretation of Decision Theory. *Ergo: An Open Access Journal of Philosophy*, **7**.
- Ullmann-Margalit, Edna, & Morgenbesser, Sidney. 1977. Picking and Choosing. *Social Research*, **44**(4), 757–785.
- van Inwagen, Peter. 1983. *An Essay on Free Will*. New York: Oxford University Press.

- van Rooij, Robert. 1999. Some analyses of pro-attitudes. *Pages 534–548 of*: de Swart, Henriette (ed), *Logic, Game Theory and Social Choice*. Tilburg University Press.
- von Fintel, Kai. 1999. NPI Licensing, Strawson Entailment, and Context Dependency. *Journal of Semantics*, **16**(2), 97–148.
- Watson, Gary. 1975. Free Agency. *Journal of Philosophy*, **72**(8), 205–20.
- Weintraub, Ruth. 2012. What Can We Learn From Buridan’s Ass? *Canadian Journal of Philosophy*, **42**(3-4), 281–301.
- Williams, Bernard. 1973. *Problems of the Self*. Cambridge [Eng.]: Cambridge University Press.
- Williams, J. Robert G. 2019. *The Metaphysics of Representation*. Oxford: Oxford University Press.
- Williamson, Timothy (ed). 1990. *Identity and Discrimination*. Cambridge, Mass., USA: Wiley-Blackwell.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.
- Zyglewicz, Tomasz. forthcoming. Kratzerian ‘Want’, Decision Theory, and Upward Entailment. *Analysis*.