Underdetermination and the Will

A dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Philosophy.

By

Kelly Whelan Heuer, M.A.

Washington, DC
September 27, 2013

Copyright 2013 by Kelly Whelan Heuer
All Rights Reserved

Underdetermination and the Will

Kelly W. Heuer, M.A.

Thesis Advisor: Maggie O. Little, Ph.D.

Abstract

Good choices are guided by the balance of reasons for and against the options at hand; bad ones made in defiance of it. This suggests a comparative standard of practical reasoning: do what you have most reason to do, on pain of irrationality. Yet many decisions are rationally underdetermined. When there is no favored alternative, we must choose by sheer will—the force of reasons cannot guide us. What does underdetermination mean for us as agents? Some see it as a threat to self-intelligibility; others as an opportunity for identity-forging choice. This dissertation analyzes both perspectives, finds them important but flawed, and thematizes the basic truths about human agency each highlights to generate a deeper understanding of the connection between agency and choice.

Acknowledgments

"It's only six years," Mark says, leaning forward in his chair. "But it's not like those six years *don't matter*!"

It's 2007, and I'm sitting in Mark Murphy's office clenching a stack of notes and a campus map, a prospective student overwhelmed by options, scrambling to balance a thousand factors to choose a graduate program. We are talking how a school's reputation is often treated as the single weightiest consideration in such deliberations, to the exclusion of all others. *But this is nonsense*, Mark tells me, and the second he says it I am shattered by its obviousness. *This is six years of your life.*

I love this moment, which has since become a favorite bit of Heuer family lore.

First, there is poetic justice in the fact that my time at Georgetown begins with underdetermined deliberation and ends with deliberation (so to speak) about underdetermination. This dissertation is a meditation on several themes crystallized in that 2007 conversation: that rational decision-making is complex; that underdetermination can take a variety of forms; that our choices inform our lives in ways that can't be undone; that underdetermined choices somehow *matter* in spite of their uncertainty—or maybe because of it.

More importantly, that moment perfectly manifests the value of the choice I did make. Six years is far too short a time to spend in a place like Georgetown, an academic community I chose because the people who comprise it are models not just for philosophical acumen but for the way that philosophy can enrich, inform, and find a place within a full human life. Among such role models, I was each year flooded with fresh reasons for humility and gratitude.

Chief among these reasons are Maggie Little, Henry Richardson, and Mark Murphy. It's a poorly-kept secret that I was doodling this dream team in the margins long before it was time to settle on a topic; I'm pleased that the fates spared me the ordeal of having to choose between that lineup and a personally compelling topic by providentially aligning them. My committee has astonished me time and again with their generosity, warmth, and brilliance. Working with three scholars I admire so much as human beings—whose strengths complimented one another so perfectly, and whose friendship enlivened not just the dissertation process, but this entire six years—was a privilege beyond all others.

For additional encouragement along the way, I have to thank big brothers Selim Berker, Dave Plunkett, Jeff Sebo, and Marcus Hedahl, who adopted me into the broader world of philosophy and showed me what it might mean to find a home there, as well as Wayne Davis, who has as department chair been unflaggingly supportive in matters large and small. The Woodrow Wilson Foundation generously funded my last year of writing with the Charlotte W. Newcombe dissertation completion fellowship, a vote of confidence I'll always treasure. (The money was nice, too.)

Most of all I am grateful to my family. My parents, Richard and Teresa Heuer, supported my love of ideas and dreams of the academy from the very beginning. I could not ask for more perfect models of temperance, discipline, hard work, and unconditional love than they have given me through the quiet example of their own lives. Amy Heuer cheered on this work from across the sea with wit, editorial prowess, sloth bear photos, and (above all) a sisterly mix of exasperation and pride in the feather-brained older sister she always knew would carry it off.

I turn at last to thank Arjun Dhillon, my found family. You never doubted it a moment, but I am not sure this could have been finished without you. It certainly wouldn't be half what it has become since you happened to it, nor would its author. Know that you are, in this as in so much else, the *sine qua non*.

❧

Table of Contents

1 Introduction.....1

1.1 Reasons & intelligibility.....2
1.2 Varieties of underdetermination.....4
1.3 A preview of the argument.....14

2 Classical rationalism.....17

2.1 The classical model.....19
2.2 Dogmatic rationalism.....29
2.3 Problems for the view.....36
2.4 The missing agent.....45
2.5 Conclusion.....47

3 Hybrid voluntarism.....50

3.1 The voluntarist turn.....53
3.2 Sheer voluntarism.....55
3.3 Problems for the view.....61
3.4 Normative voluntarism.....64
3.5 Problems for the view.....71
3.6 Conclusion.....92

4 Radical voluntarism.....95

4.1 Radicalization.....97
4.2 Heroic existentialism.....100
4.3 Problems for the view.....109
4.4 Existentialist constructivism.....126
4.5 Problems for the view.....135
4.6 Conclusion.....142

5 Conclusion.....145

5.1 Assumption one: methodological.....148
5.2 Redeeming the assumption.....162
5.3 Assumption two: conceptual.....167
5.4 Redeeming the assumption.....170
5.5 Conclusion.....185

6 Bibliography.....189

1 Introduction

The most widely accepted understanding of rational choice—so widely accepted as to be virtually a truism—holds that we ought to do what we have most reason to do. Rational choice aims at what is favored by the balance of reasons. As a metric of choiceworthiness, reasons have an essentially *comparative* function, weighing against one another for the purpose of determining what to do. Paradigmatically rational choice is guided by the principle 'do what you have most reason to do, relative to available alternatives.'

The comparative conception of rational choice has a fine philosophical pedigree as well as intuitive heft. After all, it is hard to understand what the calculus of favoring and disfavoring considerations in practical reasoning is *for* except to select the 'best' option. Yet there are cases where comparison stalls. You might face a set of options where two or more of the best alternatives are supported by equally strong, or equally good, (sets of) reasons. The balance zeroes out. Or perhaps you face incomparable options: the

alternatives cannot be rank-ordered, but neither is it true that two or more are exactly equal, perhaps because they realize incommensurable values. Or perhaps there is no stable fact of the matter about which option is most supported by the balance of reasons, because of some shifty variable in the calculus. Call these instances of *rational underdetermination*: there is no option univocally favored by the balance of reasons.

In each case, the comparative principle runs into trouble. Understood formally, as a criterion of rational choice, it gives no answer to the question of whether someone ought to have chosen otherwise than she did. Paradigmatically rational choice is impossible. Understood substantively, as describing what guides ideal rational deliberation, it gives no concrete guidance, since 'do what you have most reason to do' no longer admits of a *de re* reading. The merits of the options make each worth doing, in some sense, but they do not make any single option 'the thing to do' when comparison cannot find a victor.

1.1 Reasons & intelligibility

Theses about the standards for rational choice, the nature of rational deliberation, the value of rational action, and so on, often co-travel—linked by the power of reasons to confer *intelligibility* in choice. A reason makes sense of the option it favors by giving some explanation or justification for its selection. Rational comparison offers forward-looking rational guidance as well as *ex post facto* justification, secures self-intelligibility and features in the explanations we give others for why we acted as we did. Reasons—prospective and retrospective, introspective and interpersonal—are base currency in the economy of meaning that makes our lives comprehensible and (if we are lucky) valuable. Our responses

to the balance of reasons as we saw it anchor the stories we tell about how and why we moved in the world as we did. When comparison fails, we lose the coin that buys the meaning from which our lives are built.

Or so one venerable tradition in philosophy would have us believe. The **classical** conception of agency unites the comparative principle with thoroughgoing rationalism about human agency: agency is manifested in all and only those choices that select an option favored by the balance of reasons. Irrational or arational action is, on some level, a failure of agency.[1] Without rational comparison, there can be no intelligible choice; if most or even many of our choices fail to meet the comparativist standard, life verges on the absurd.

The classical view is predominate in the history of Western philosophy, and continues to loom large in contemporary theories of ethics, action, and practical reason. But it has not arrived on the contemporary scene unchallenged. As an account of meaningful human agency, it has so far withstood a series of **voluntarist** challenges, which identify acts of the *will* as the locus of meaning and value in human action. On this view, willful choice may be unconstrained by the balance of reasons, and so the voluntarist treatment of underdetermination must find some notion of intelligibility outside the parameters of comparative justification. On this view, to be an agent is, sometimes, to act arationally or even irrationally: for whims, for no reason, even for bad reasons.[2] Rather than a threat to

---

[1] Davidson 1970 is a canonical modern statement of the view; see also Hare 1952, 1963; Mele 1995, 2003; Audi 1990, 1993; Bratman 1979.

[2] Frankfurt 1987, 1998a, 1998b, 2004; Slote 1989, 2001; Velleman 1992; Buss 1999; Millgram 1997; Taylor 1997; Tenenbaum 1999; Arpaly 2003, Chang 2004, 2009. See also Schroeder 2009, pp. 189–91.

self-governance or self-understanding, the voluntarist sees underdetermination as a site of identity-forging choice and the source of a valuable form of human freedom. Only when our choices fly free from rational constraint do our choices (and our lives) take a form that is truly 'up to us.'

Members of these groups are not often explicit about the commitments that motivate them, especially contemporary theorists. But notice how dramatically these conceptions of what rational underdetermination *means* differ from one another: is it a threat to meaningful human agency or a condition on its very possibility? I take this philosophical disagreement to be the best evidence that the phenomenon of underdetermination is worth our attention. We turn in the next section to a catalogue of the ways in which reasons can underdetermine choice, not with an eye to showing that underdetermination is a common phenomenon—though I happen to think it is—but rather as a first step toward understanding the ways that theories of rational agency might be upset by it.

1.2 Varieties of underdetermination

Underdetermination occurs when the normative pressures of some domain fail to deliver a definitive answer to the question of what ought to be done or believed, from the perspective of that domain. For some domain **D**, there is no answer to the question: 'what do I have most **d**-reason to do?' There are, I suggest, four major varieties of domain-specific normative underdetermination. Each has bearing on the question that interests us here: the summed-domain underdetermination of the all-things-considered rational perspective.

Inapplicability

When a domain **D** does not extend to cover the choice one faces, there can be no **d**-reasons for or against the options at hand. Plausibly, for instance, the choice of which toothpaste I use to brush my teeth this morning is underdetermined by legal normativity: there simply are no legal reasons for or against the options at hand. If I want guidance in my choice, I must turn elsewhere: to prudential reasons, say, which tell me to use the one that makes my mouth feel freshest, or perhaps to moral reasons, which tell me to use the tube that belongs to me and not the tube that belongs to my sister.

Inapplicability is an historically important form of underdetermination because of the straightforward way it seems to secure freedom of meaningful choice: when the reasons of some normative domain of a higher lexicographic order fall silent (legal, moral, religious), those from less 'important' domains (social, prudential, aesthetic) are permitted to have their say.[3] One way of understanding political freedom is as widespread legal underdetermination—there are matters in one's life about which government institutions simply have nothing to say—and many contemporary arguments against 'totalizing' moral views like consequentialism, which purport *always* to be in the business of offering you moral reasons for choice, use precisely this analogy to argue that such views obviate the

[3] This grouping of 'important' versus 'less-important' classes of reasons for action is just meant to gesture at a rough consensus view; if you disagree that, e.g., there are such things as distinctly religious reasons, or that moral reasons are in some sense structurally distinct from other sorts of reasons, simply omit or reshuffle the given categorizations.

possibility of acting for reasons "of one's own," pursuing plans and projects whose choiceworthiness is not moral but personal.[4]

Can there be cases where *all* normative systems fall silent? If one takes the view that reasons only get to be called reasons when they achieve a certain threshold of favoring force, many trivial choices—which side of my mouth to begin brushing first, to return to our example above, or to exercise 0.0001 psi of pressure more or less when I start brushing—are underdetermined because there are quite literally no reasons at all bearing on the choice. Whatever considerations might advocate right over left, or harder over softer, are so inconsequential they do not deserve the title 'reason.' If so, then inapplicability is a kind of underdetermination that can characterize the all-things-considered perspective of practical reasoning.

Actually, I think the threshold view is a theoretical mistake, which fashions a theory of what considerations must do to qualify for their status as 'reasons' out of the standards for what a reason must be if it is to be worthwhile for us, in ordinary circumstances, to expend deliberative resources on considering its weight in our choice-making.[5] Considerations that are suited to play a role in explaining, guiding, or justifying *even very trivial* choices have the same structural features as more robust reasons, even if they do not actually play that role in the deliberative economy of healthy agents, since only the obsessive and the neurotic are likely to fixate on such minutia when there are other matters to attend to.

---

[4] See, e.g., Williams 1993, Rawls 1971; but *cf*. Herman 1993, who argues for moral choice as a form of self-expression and, blurring the moral–social distinction in the other direction, Stohr 2006.

[5] This is a view shared by, among others, Schroeder 2008.

But laying this debate aside, there is a more interesting form of inapplicability that looms at the very other end of the spectrum: the idea that it is not our most trivial but our most consequential or foundational choices that might fall outside the bounds of rational assessment from any domain. This is a prospect famously explored in Søren Kierkegaard's *Either/Or*, which describes two fundamentally different approaches to life: the 'aesthetic' and the 'ethical,' each of which provides an internally self-consistent set of normative standards and ideals to guide choice. The normative systems are, moreover, self-ratifying: the ethicist (B) can offer no reason for the aesthete (A) to switch teams that will be accepted unless A has already committed himself to an ethical way of life, and vice versa.[6] The choice between normative systems, obliquely presented to the reader by Kierkegaard's pseudonymous editor, Victor Eremita, is fundamentally arbitrary:

kierkegaard's editor You must choose between competing, self-ratifying normative systems.

There can be no reasons that do not come from within one of these systems for adoption or rejection of either. It is a unique kind of inapplicability: no reasons apply to your choice because the choice itself, by definition, places you outside the bounds of any of the normative domains which might otherwise speak to the choice. Any considerations in

[6] Kierkegaard 1843/1998. Actually, Kierkegaard followed the publication of *Either/Or* just a few months later with a work in his own name, *Two Upbuilding Discourses* (1843), which seems to offer a resolution to the aporetic ending to A & B's dialogue: the adoption of a religious perspective, which supersedes both the aesthetic and the ethical as a normative framework, but must be adopted on the grounds of faith not reason.

favor of one option over the other speak with a force that cannot touch you until you endorse their validity: but that is precisely what the choice is meant to settle.[7]

Indifference

A more familiar sort of rational underdetermination—perhaps the most familiar of all—is the strict "indifference of alternatives" faced by Buridan's ass in the famous paradox.[8]

buridan's ass	Standing equidistant between two identical bales of hay, you are hungry and want a bite to eat.

Traditionally, the donkey's dilemma is that withholding choice is both the route dictated by the standard conception of rational justification and clearly a foolish thing to do, all things considered. Either your action, when you succumb to hunger and simply 'pick' (rather than 'choose') one bale of hay over the other, lacks rational determination; or you stand between the bales until your legs give out beneath you from hunger, never having acted against reason but much worse off than the donkey who just flipped a coin (so to speak) and started munching. Like inapplicability, this form of underdetermination can occur both within and across domains: the donkey's difficulty might seem to stem either from an equivalence of normative forces in the prudential domain, or in the summed-domain perspective of practical reason.

---

[7] This is, of course, the existentialist interpretation of Kierkegaard's work—not all commentators agree that the message of *Either/Or*'s aporia has to do with the radical rational underdetermination of our choice of life values. See, e.g., Hannay and Marino 1997; Davenport and Rudd 2001.

[8] Although this nomenclature appears to attribute authorship of the paradox to 14th century philosopher Jean Buridan, we have no record of his authoring the example. Indeed, it seems likely that the ass bears Buridan's name because it generates a *reductio ad absurdum* of Buridan's own theory of the will, though it's also been conjectured that he used the example in oral disputation. More on this in §**2**.

A variant on this dilemma predicates indifference not on the indistinguishability of alternatives, but instead on the equality of preference an agent has for them. This paradox makes an appearance as early as 350 B.C.E., in Aristotle's *De Caelo*.[9]

aristotle's dog Standing equidistant between equal volumes of food and water, you are as hungry as you are thirsty, and darn thirsty at that.

The plight of Aristotle's dog[10] is typically taken to mirror that of Buridan's ass: the donkey has no reason to choose one bale of hay over the other, but good reason to eat either rather than remain hungry; the dog has no reason to eat rather than drink, but good reason to do each, since she's equally hungry and thirsty.

The case is interesting because it opens up a new range of considerations. Some of those who've tackled the traditional ass's paradox have actually denied that such cases are possible since the world will nearly always present us with some empirical difference between alternatives that could, in theory, serve to justify choice.[11] Modern incarnations of the ass's problem, such as the shopper facing a row of identical cans of soup needing to select only one for dinner, might also be susceptible to this treatment—probably one can is closer to hand than all the others, if nothing else—though §2 will suggest that this dogmatic treatment of indifferent choice is often unsuccessful.

---

[9] Aristotle, *De Caelo*, II 13 295 b 24.

[10] The original text clearly identifies the protagonist as a man, but Jean Buridan himself (in an unpublished commentary on *De Caelo*) makes the protagonist a dog instead, and I've adopted his version on grounds of cuteness.

[11] Gottfried Wilhelm Leibniz is perhaps best known for this claim, but see also the storied history of disputes over rational indifference surveyed in §2 of this work, in which many (including Thomas Aquinas, Moses Maimonides, and al-Ghazâlî, to name just three) endorse a similar position.

The dog's situation moves past the comparison of identical alternatives to ones which are qualitatively distinct, with preference alone contributing the equalizing force. Any hungry donkey or shopper will face the challenge above; only someone with both hunger and thirst will even be in a position of needing to compare the hunger-satisfying and thirst-slaking properties of two options. Put another way, the case of Aristotle's dog presupposes the comparability of different alternatives. In this case, it seems fairly straightforward to find some unit of commensuration—degrees of physical satisfaction, perhaps—that enables comparison of food and drink. But an entire literature on incommensurability and comparative choice challenges precisely this presupposition, and opens up a new way in which guidance by the balance of reasons might be blocked.

Incomparability

A choice may be rationally underdetermined if the options at hand are incomparable: if it cannot be said of two alternatives a and b that [i] a is more choiceworthy than b, [ii] b is more choiceworthy than a, or [iii] a and b are equally choiceworthy. Guidance by the balance of reasons is impossible here because there is no balance, properly speaking. It has sometimes been said of such choices that they manifest incommensurable values: values that cannot be measured on a common scale.[12] An alternative explanation of the phenomenon focuses less on the absence of a suitable *comensurans* and instead diagnoses the problem as one of vagueness, whether of the values exhibited in each alternative[13] or of the

[12] For an overview of the state of debate over incommensurability, incomparability, and practical reasoning, see Chang 1997.

[13] This is the explanation favored by, e.g., Raz, 1988, 1997.

"better than" relation itself.[14] As a domain-specific phenomenon, incomparability has often been thought to characterize moral or aesthetic choices, where the considerations that favor various options either are or are closely related to concrete or distinct kinds of value: the expressiveness versus technical acuity of a cello solo, the moral value of friendship against the impartial demands of justice, the haecceity of this individual whose life I can save by sacrificing the life of another.[15]

Whatever metaphysics of value underwrites the phenomenon, incomparability is thought to be manifested in cases like the one described by Jean-Paul Sartre in "Existentialism is a Humanism,"[16] of a young student in Nazi-occupied France torn between two courses of action:

sartre's student	You must decide whether to join the French Resistance or to stay home to take care of your ailing mother.

[14] Broome 1997. Griffin (1997) also argues that incomparability involves vagueness, but his account differs in some details.

[15] Aristotle, for instance, remarks in *Nicomachean Ethics* that "[t]he spheres of what is noble and what is just … admit of a good deal of diversity and variation" and because "goods vary in this way as well," "we should be content … to demonstrate the truth sketchily and in outline," since after all "the same degree of precision is not to be sought in all discussions, any more than in works of craftsmanship" (1094b, 12–18). For contemporary discussion of this idea see Railton 1992, esp. pp. 704ff; Sainsbury 1989; Rosen and Smith 2004; Hyde 2008; Schiffer 2010. For an example of domain-specific (aesthetic) underdetermination percolating up to the level of rational underdetermination, see Chang 1997, pp. 14ff.

[16] Sartre 1946.

Whether we understand this incomparability in terms of vagueness or in terms of incommensurability,[17] the thought is that it seems impossible to find a determinate all-things-considered judgment about what it is best to do. The options are so dissimilar that a useful comparison is beyond reach. To put it somewhat fancifully, the reasons that speak in favor of each options cannot speak to one another—or at least speak over one another to the person deliberating. The normative outcome is not so much silence as it is white noise.

Indeterminacy

I shall have more to say about incomparability in subsequent chapters. For now, though, let's consider a different kind of underdetermination, which characterizes choices that involve infinite values or vague predicates. The cases of **buridan's ass**, **aristotle's dog**, and **sartre's student** all involve temporally-bounded choice: one must choose, at time *t*, from some finite set of options. But what if one's choice set were infinite? No problem, so long as the set contains some maximal element: the infinity of inferior options need not trouble you. Issues only arise when a choice set presents more than one optimal element, but no maximal element—as is the case with a series of paradoxes regarding choice over time.

A paradigm instance is the following, taken up by, inter alia, Gottfreid Wilhelm Leibniz:

---

[17] Sartre himself appears to endorse both and neither: "If values are uncertain, if they are still too abstract to determine the particular, concrete case under consideration, nothing remains but to trust in our instincts. That is what this young man tried to do… But how does one estimate the strength of a feeling? […] I can only estimate the strength of this affection if I have performed an action by which it is defined and ratified." (1946). And so the deeper explanation on offer—explored in §**4**, and foreshadowed in §**1.2.4** of this section—has to do with the fundamental impossibility of understanding our choices prospectively rather than retrospectively. Only choice itself can fix the value of some inputs.

leibniz's god You have an infinite number of possible worlds you might choose to create, and an infinite number of instants in which to create them.

Famously, Leibniz held the existence of a world to demonstrate not just that there must have been an optimal time for the world to have been created, but also that the world that was created—the one in which we live—is the best of all possible worlds, since God's "supreme wisdom…cannot but have chosen the best" and "if there were not the best (*optimum*) among all possible worlds, God would not have produced any" at all.[18] Others, however, have argued for a resolution to the problem that appeals to the power of God's will to create a normative truth of the matter about when and how to create the world, or to God's omnipotence as extending to a kind of freedom even from the principles of rational choice themselves.[19]

At any rate, the paradox is recapitulated in non-theological examples like that of an eternally-maturing charitable trust (at any time t you could help a great number of people by terminating the trust and giving the balance to charity, yet you could help more people by terminating it at $t + 1$, when still more interest has been generated) or, more fancifully, an ever-improving bottle of wine (given an infinite lifespan and a bottle of ever-improving wine, it will be irrational at any t to open the bottle rather than waiting until later to open

[18] Leibniz *Theodicy* §8, p. 128.

[19] This paradox dates at least to theological debates of early antiquity, e.g. the work of al-Ghazâlî and Averroes, and others among the Arab and Jewish theologians who wanted to defend the Abrahamic notion that the world was created rather than eternal against the Greek philosophers. For more see Averroes 1954, which contains extensive quotation of and commentary on Ghazâlî's challenge. Related paradoxes in which the unique features of God-as-rational-choice-maker also cause some trouble can be found in Rowe 1994, Adams 1999, Wielenberg 2004.

it, yet more irrational still not to savor such extraordinary wine at some point).[20] In such cases, the choice involves some constantly shifting variable in the rational calculus. Though finite beings are perhaps rarely presented with infinite choice sets, vague, variant, or unfixed inputs to the rational calculus might render a normative verdict indeterminate for mere mortals as well.

1.3 A preview of the argument

So the kind of normative silence that underdetermination presents us with is multiform: it can arise from substantive or merely formal features of the case at hand, can terminate or simply block deliberation altogether, can involve a normative void or a kind of cacophony. The practical problem that underdetermination presents—what to do when there's no 'thing to do'—exists in all cases, but the variability of the normative materials at hand in different cases might help explain the radically different roles granted to arational volition in the theories of human agency canvassed in §1.1. If the **classicists** tend to concern themselves with the structural underdetermination of indifference and indeterminacy, as I suggest, while certain **voluntarists** look to the normatively complex phenomena of incomparability and inapplicability, we are some part of the way to understanding the classical rejection of arbitrariness and the voluntarist embrace of it.

However, the roots of this divide run deeper than any superficial treatment of underdetermination, to the notion of *intelligibility* that seems so naturally to tie what is meaningful and valuable in human choice with a comparative rational standard. The

[20] Pollock 1983 is the author of the ever-better wine paradox; Landesman 1995 the key contemporary discussion of the charitable trust paradox.

dialectic centers around an apparently inescapable tension between the classical notion that intelligible choice requires comparative justification, and the voluntarist idea that comparative rational constraint may blot out the spontaneity, creativity, or—most crucially—self-determination which are the *true* heart of agency. If all or even only the most important exercises of our agency involve acting as a conduit for the translation of rational pressure into action, human life seems grimly mechanistic, protests the voluntarist. Yet against this we can hear the classical bewilderment: why suppose that our most valuable exercises of agency take place precisely where our ways of understanding action run out?

Much theoretical disagreement in the contemporary literature about rational underdetermination is, I will suggest, driven by largely unspoken allegiances to one of these two orientations. Contemporary theorizing about underdetermination will go better if we are clearer, in the first instance, about what forms of underdetermination are really under discussion, a clarification with which the typology in §1.2 should help. But it will also go better if we are more honest—or perhaps just more self-aware—about the implicit models of meaningful choice that shape many of our downstream judgments about when underdetermined choice is possible. This awareness can be inculcated by exploring key themes that emerge from the historical dialectic between classicism and voluntarism, and by increasing our sensitivity to what I will suggest is the essential intractability of the debate.

This dissertation is, above all else, a map. The next three chapters sketch divergent positions in conceptual space, rendered contiguous by their commitment to a comparativist rational standard and differentiated by their account of the centrality of

willing to meaningful human choice. The map itself is not an argument, though there will be many arguments about the plausibility or stability of certain claims along the way. (We are philosophers, after all, and not explorers only.) What the map offers is information about the cost of taking up residence in any of the regions under exploration: how much can be gained, and how much must be given up, when we commit to one theory of rational agency over another?

Our exploration is driven by the conviction that responsible philosophizing about rational agency involves facing our own implicit allegiance to some territory on the map. It ventures through some creative attempts to carve out a habitable space between **classical rationalism** and **radical voluntarism**—a class of views I call **hybrid voluntarism**—but finds them either unstable or untenable. The hybrid views are found to tumble back into classicism, slip toward full radicalization, or disintegrate under the force of these equal and opposite gravitational pulls. There are, in fact, uninhabitable regions of the map. And with a twist of something like irony, we find that the habitable but inhospitable regions—the classical and radical theories of rational agency—can in the end do no more than offer themselves for adoption in what is perhaps the ultimate underdetermined choice.

2 Classical rationalism

Perhaps the most remarkable thing about underdetermination is that for much of the history of Western philosophy, the general view was that there is in fact no such thing. This is not because the phenomenon was overlooked: as §1 makes clear, discussions of indifference in choice date back at least to the dilemma of **aristotle's dog**, and **buridan's ass** was a matter of no little theological debate in the medieval, Renaissance, and early modern periods.[21] Rather, the bare possibility of normative uncertainty was ruled out by classical theses about the rational organization of a teleologically- or divinely-ordered world, and the resultant nature of the choices made by rational agents within them.

In this chapter, I explore what I call the 'classical model' of rational choice, which links the evaluative standards for reasoned choice with the constitutive standards for rational human

[21] Indeed, quite possibly discussion begins earlier than that, depending on how you draw your lines —see Rescher 1960 for discussion of pre-Socratic discussions of indifference and indeterminacy in the balance of physical rather than rational forces.

action and offers a strong comparative reading of the notion that rational action is guided by the balance of reasons. The classical model, structured by theses about the nature of human choice such as the *guise of the good* and the *principle of sufficient reason*, offers a tidy and self-consistent account of rational choice, but at a high cost. Its strongest versions, which rely on metaphysically robust renderings of the Good and its relationship to reasons for action, are forced to reject the very possibility of arational choice and can offer no real explanation for irrational choice either—in plain defiance of our lived experience. Further, as contemporary critics pointed out, these views seem also to rule out the possibility of free will—or at least of any meaningful notion of self-determination. Weaker versions, which offer metaphysically neutral versions of the classical dicta about rational choice, escape these two problems but fall prey to a more subtle variant of the free-will concern: that self-determination is once again ruled out by the totalizing conception of rational determination entailed by the conflation of rational and constitutive standards for choice.

These last two concerns can both be traced to what I describe as the 'hydraulic' model of willing entailed by the classical model. In a slogan, there seem to be problems with free will because for the classical rationalists there is nothing for the will—understood as an active principle in moving agents from deliberation to action—to do or to be; it adds nothing to a system that already contains all the elements required to explain and justify human choice. The classical theory succeeds in explaining the sense in which we take ourselves to be *rational*, and fails to explain the sense in which we take ourselves to be *agents*, because it essentially reduces the latter to the former.

2.1 The classical model

The idea that one cannot intelligibly act in ways one does not take to be best is one of the oldest in philosophy: traceable at least to Socrates, who famously argued that an agent cannot pursue a course known to him to be worse than some alternative. Plato, and Aristotle after him, both held versions of the view, but the theory offered by Thomas Aquinas many centuries later has had perhaps the most influence on contemporary theories of practical reason. The so-called "guise of the good" thesis—which holds that all action is undertaken *sub specie boni*, under some idea of it as good—can be construed variously as a doctrine about human nature, a theory of religious progress, an analysis of the nature of desire, or a definition of rational action.

It is, of course, this last which concerns us here. Recall our working account: rational choices are guided by the balance of reasons. Classical rationalism, as I define it, offers a *strong comparative* reading of the notion of a 'balance' of reasons: it is rational to do what you have most reason to do, and irrational to do anything else. These rational standards also serve as constitutive standards for intelligible action: recognizably human choice is guided by deliberation about what one has most reason to do; any action undertaken in the absence or in defiance of such standards is not just irrational but in a certain sense incomprehensible. On some views, indeed, it is literally impossible.

Throughout its long history, the guise of the good thesis has in fact had many guises. It is, at root, an account of the relationship between some directed state or activity in an agent (desire, wish, inclination, intention, and so on) and the object of that state (apparent or actual good, self-interested or moral good, immediate good or long-term good, and so on).

Thus it can be multiply specified, along at least two axes:

i · nature of state (*conative, cognitive, volitional; both conative and cognitive; etc.*);

ii · nature of object (*objective, subjective; individual, collective; prima facie, all-things-considered; etc.*).

Perhaps the most robust statement of the thesis incorporates *all* such positively-directed states: desires, inclinations, intentions, choices, and actions are each, as such, directed at the good. What sort of good? We may differentiate between what an agent merely takes to be good and what is objectively good, then again among the various ways an option might be valued: as individually or collectively good, as pleasant or as worthwhile, as *prima facie* good or good all-things-considered, as instrumentally or intrinsically good, and so on.

The ancients generally affirmed some broadly inclusive permutation of the above, offered as a theory not just of desire (*orexis*) and appetite (*epithumia*), but also of distinctively rational conative states such as wish (*boulêsis*) and decision (*prohairesis*). Discussion typically identifies action and choice as manifestations of these internal modes.[22] Whether the direction of such states is toward the Good itself or instead toward those things taken to be good by the agent herself was a matter of some debate, with Plato unsurprisingly affirming the former, Aristotle the latter, and Socrates, as ever, enigmatic.

[22] See, e.g., *Meno* 77a078c: "No one goes willingly toward the bad or what he believes to be bad; neither is it in human nature, so it seems, to want to go toward what one believes to be bad instead of to the good." Similar sentiments can be found in the *Gorgias* (466a–468e) and the *Symposium* (205d, 205e7–206a1).

What's notable is that all three, and others who later followed them, seem to have opted for a strictly comparative notion of 'goodness' as the appropriate object of these inclinations. Here is Socrates in Plato's *Protagoras*, for instance:

> …no one who knows or believes there is something else better than what he is doing, something possible, will go on doing what he had been doing when he could be doing what is better.[23]

And interestingly, the very discussion of balanced rational forces that gave us the example of **aristotle's dog** was cited by Aristotle by way of making an analogy with the balanced physical forces that, on some ancient views, hold the world in place.[24] The goodness of the objects of appropriate human inclinations is not a matter of meeting some threshold, nor is it an essentially *prima facie* notion. It is not enough that the object has something to speak in its favor—some reason to pursue or desire it—it must be judged superior to other available options.

The principle of choice that characterizes the classical discussion, then, is not so much 'the guise of the good' but 'the guise of the best,' and it casts a long shadow in the history of philosophy. Early versions of the principle were often supported by some broader thesis about the nature of the world within which willing and choice take place: as teleologically structured by an all-encompassing *logos* or rational world soul, as the Stoics had it; ordered by a *demiurgos* with intrinsically compelling Forms to guide and orient human choice, as in Plato's vision; or providentially organized by an omnipotent, omniscient, and omni-benevolent Creator, as for the Arabic, Jewish, and Christian theologians of late antiquity

---

[23] *Protagoras* 358b7–c1.

[24] *De Caelo* II 13, trans W.K.C. Guthrie in the Loeb Classical Library series.

and the middle ages. There were some ripples in the fabric of this conception, as we shall soon see, but it is worth noting that in many cases the classicists blend rational with constitutive standards because they view them as mutually entailed by some further, independently motivated thesis about the nature of causation or active principles in the natural world.

This is perhaps nowhere more explicit than in the work of early modern inheritors of the classical mantle. Here the guise of the best thesis is given its broadest extension, and its implications for human choice are most ruthlessly pursued. Gottfreid Wilhelm Leibniz, who coined the term 'Principle of Sufficient Reason' (hereafter: **PSR**) to refer to the dictum that "nothing happens without a reason," alternately refers to it as 'the Principle of the Best' precisely because of the tight relationship between goodness, Reason, and causation that characterizes the classical picture.[25] Though Leibniz coined the term, the Principle of the Best is also a centerpiece of fellow rationalist Baruch Spinoza's slightly earlier work.[26] We turn now to these two men for a distillation of the classical theories of rational choice and free agency.

Like the guise of the good thesis, the **PSR** comes in different flavors, depending on:

i · modal strength (*necessary, contingent*);
ii · scope or extension (*facts, actual events, possible events, etc.*).

For our purposes, the question of whether we understand the **PSR** as necessarily governing or contingently characterizing facts or events is less interesting than that of how to

[25] Leibniz and Clarke 1715; Leibniz, *New Essays Concerning Human Understanding*.

[26] It is also a close philosophical cousin of the earlier Cartesian 'Principle of Sufficient Cause' used in his proof for the existence of God in the third of his *Meditations on First Philosophy*. Descartes 1641.

understand the spirit that motivates it. Already implicit in the second axiom of the first part of Spinoza's *Ethics*,[27] the principle is given canonical expression a bit later:

> For each thing there must be assigned a cause or reason, both for its existence and for its non-existence.[28]

Spinoza's driving concern is the *explicability* of the world in which he finds himself, from the phenomena of social and political life to the movement of the heavens. In the view of early modern rationalists like Descartes, Spinoza, and Leibniz, the universe itself must be intelligible from top to bottom, the rational activity of its human inhabitants no less than the causally-determined movement of its physical parts. Intelligibility is rendered in the language of rational 'conceivability'—reasons–explanations—and with the possible exception of God, no entity can be self-caused or self-explained.[29, 30] Every phenomenon is backed by reason: not just *a* reason, but *sufficient* reason, where 'sufficiency' is given a strictly comparative reading. Self-explaining facts, self-caused activity, essentially unpredictable or random phenomena, even the slight deviations from 'bestness' suggested by the children

---

[27] "What cannot be conceived through itself, must be conceived through another." *Ethics*, EIa2.

[28] *Ethics,* EIpIId2.

[29] On conceivability and reasons-explanations, see the *Ethics,* EIa2, EIa4, EIp10s, EIp14d; see also Della Rocca 2008 for a much broader explication of this general conflation of conceivability and causation, and their ultimate reduction to some notion of intelligibility or explicability itself.

[30] On the notion that God ('true thought') might be self-explained, see Spinoza's somewhat enigmatic remark in his *Treatise on the Emendation of the Intellect:* "that Thought is also called true which involves objectively the essence of some principle that does not have a cause, and is known through itself and in itself" (II.26, pp. 33–34). This is Spinoza's commentary on Descartes, and might refer to the Cartesian view that while God's existence is uncaused, there good reason for this exception, and so there is a reason but not a cause for God's existence Melamed and Lin 2011, §**2**. Leibniz was far more comfortable countenancing the existence of self-explaining entities like God, as we shall see, but the fact that such powers were not extended to the creaturely realm means that the Principle of the Best holds firm sway in human affairs in any case.

and fools above—all lie outside the boundaries circumscribed by the PSR and therefore have no secure place in the rationalist worldview.

Most early discussions of the collision between classical rationalism and the possibility of underdetermination are oriented around examples which feature the strict **indifference** of alternatives—as is the case for "the Ass of Buridanus" and the dog or man in Aristotle's *De Caelo*—or the mathematical **indeterminacy** which seems to characterize divine choices combining omniscience and supreme rationality with an infinity of options.

Recall the structure of such cases:

buridan's ass	Standing equidistant between two identical bales of hay, you are hungry and want a bite to eat.
aristotle's dog	Standing equidistant between equal volumes of food and water, you are as hungry as you are thirsty, and darn thirsty at that.
leibniz's god	You have an infinite number of possible worlds you might choose to create, and an infinite number of instants in which to create them.

The puzzle of the **ass** is driven by the supposition that qualitatively identical options necessarily offer equally compelling reasons for choice; so too the presumably commensurable physical desires of the **dog**; so too the infinity of indistinguishable instants in which the world might have been created by a perfectly rational **god**. Without some disequilibrium in the balance of reasons, how can choice proceed?

Discussions often presumed an analogue with the case of an equilibrium of physical forces, of the kind often cited in ancient discussions to explain the stability of certain bodies or elements in the terrestrial order. To take just two instances: sixth-century Aristotelian commentator Simplicius offers the following analysis of the case of **aristotle's dog**:

> The sophists say that if a hair composed of similar parts is strongly stretched and the tension is identical throughout the whole, it would not break. For why would it break in this part rather than that, since the hair is identical in all its parts and the tension is identical? Analogously also in the case of a man who is exceedingly hungry and thirsty, and identically so in both, and identically lacking in food and drink, and for this reason identically motivated. Necessarily, they say, this man remains at rest, being moved to neither alternative. For why should he move to this one first, but not that, [since] his need, and thus his motivation, is identical?[31]

Centuries later, Leibniz would himself hearken back to just such a physical analogue, citing the Archimedean dictum that "Equal weights at equal distances are in equilibrium, and equal weights at unequal distances are not in equilibrium but incline towards the weight which is at the greater distance" as a special case of his favored Principle of Sufficient Reason.[32] Nothing happens without a reason in a rationally-ordered universe, and the definition of 'reason' extends from physical forces to normative pressures, covering everything in between.[33]

---

[31] *Commentaria in Aistotelem Graecia*, vol. 7, *Simplicii in Aristotelis de Caelo Commentaria*, ed. I.L. Heilberg (Berlin: Royal Prussian Academy, 1894), pp. 533–534, as translated and quoted by Rescher 1960.

[32] Leibniz to Clarke, in Leibniz 1989, p. 321; see also his "First Truths," p. 31, *ibid.*

[33] Compare Spinoza's reductive naturalism: as he famously insists in the preface to *Ethics* III, human beings should not be considered "in nature as a dominion within a dominion [*imperium in imperio*]," as was then (and in many ways still is) common philosophical practice. We are not a special subclass of entities in the natural world to which unique laws or forms of explanation apply, but one whose activities are regulated and explained by those which govern all of nature. Later in the same preface, Spinoza insists: "Nature is always the same, and its virtue and power of acting are everywhere one and the same, that is, the laws and rules of Nature, according to which all things happen, and change from one form to another, are always and everywhere the same. So the way of understanding the nature of anything, of whatever kind, must also be the same…." For discussion, see further: Della Rocca 2008; Garrett 2003, esp. pp. 315–316; McDonough 2011, pp. 14, 25–26 (page references are to manuscript version).

The same pattern characterizes the theological puzzles which so worried theologians of late antiquity like Abu Hamed Muhammad ibn Muhammad al-Ghazâlî and modern rationalists like Leibniz. God is perfectly rational, and so could not choose anything other than the best. There would be something unsettling about the idea that any of God's choices—*especially*, perhaps, really important ones like when to create the world—were arbitrary. And yet if the world was in fact created, rather than eternal, surely there must have been some instant from an infinite and qualitatively homogenous array of empty instants from which to choose to create the world.[34] This is true whether we conceive God to exist within time or outside of it. Leibniz is insistent:

> God must needs have chosen the best, since he does nothing without acting in accordance with supreme reason.[35]

He takes the fact that a world exists to be evidence not just for the belief that there must have been an optimum time for its creation, but also that the one created must be the best of all possible worlds:

> Now [God's] supreme wisdom, united to a goodness that is no less infinite, cannot but have chosen the best. For as a lesser evil is a kind of good, so even a lesser

[34] Actually, the issue is slightly more complicated than this. A number of theologians, some directly motived by paradoxes like this one, have affirmed the eternality *and* the createdness of the world, denying that there is any tension between the two characterizations—see, e.g., the account offered by St. Augustine of such contemporaries: "for as, if the foot were always in the dust from eternity, there would always be a footprint which without doubt was caused by him who trod on it, so also the world always was, because its Maker always existed" (*De Civitate Dei* xi, 31). Aquinas notes this passage, calling the view there "barely intelligible," later suggesting that while natural reason tends towards affirming the eternality of the world, the words of Scripture suggest a beginning in time, a notion which must be accepted as an article of faith (*Summa Theologica* Ia, q. 46, a. 2). At any rate, this added nuance impugns neither the structure of the paradox itself nor the rationale for Al-Ghazâlî's response to it. I thank Mark Murphy for helpful discussion of this point.

[35] Leibniz, *Theodicy* p. 128, §8.

> good is a kind of evil if it stands in the way of a greater good; and there would be something to correct in the actions of God if it were possible to do better. As in mathematics, when there is no maximum nor minimum, in short nothing distinguished, everything is done equally, or when that is not possible nothing at all is done: so it may be said likewise in respect of perfect wisdom, which is no less orderly than mathematics, that if there were not the best (*optimum*) among all possible worlds, God would not have produced any.[36]

And so Leibniz's own solution to the divine puzzle I have given his name was, essentially, to deny its possibility.

Earlier scholars, though, were less content with dogmatic assertions of this kind. Here is al-Ghazâlî, worried in the 12th century by the heresy of latter-day Aristotelians:

> How will you defend yourselves, theologians, against the philosophers, when they … [claim] that times [for creating the world] are equivalent so far as the possibility that the Divine Will should attach itself to them is concerned…?[37]

Arabic Aristotelians like Avicenna and Averroes thought that this paradox suggested that the world was eternal and therefore uncreated, but al-Ghazâlî was not content to abandon the orthodox Muslim belief in the createdness of the world. He agreed that it would be deeply problematic, if not incoherent, to assert that in creating the world God had acted for no reason. But unlike Leibniz, he was not content to simply assert that there must have been *some* way of discerning the extant superiority of one of the infinity of options to all others. So he argued instead that the divine will, unlike those of humans or animals, had the power to effect a shift in the normative balance that characterizes any choice. It is part of God's omnipotence, al-Ghazâlî insisted, that even conceptual or normative truths may

---

[36] Leibniz, *Theodicy* p. 128, §8.

[37] From *The Incoherence of the Philosophers*, as quoted in Averroes 1954, vol. 1, p. 18.

be bent to the divine will. Whereas **aristotle's dog** and **buridan's ass** might be doomed to starve, **leibniz's god** would face no such problem, and could, without impugning a single divine perfection, make an arbitrary choice into a rational one by sheer force of will.[38]

Al-Ghazâlî's move is interesting because it identifies the divine will as an *active principle* in choice, which enables God's rational action by in some way altering the parameters of the choice at hand. While the Arab Mu'tazilites and Ash'arites generally rejected the notion that merely human willing had this power, certain later Christian theologians would suggest that the power to break normative ties or resolve normative indeterminacy was a power granted to humans as well. In order for freedom of the will—important in Christian theology for a number of reasons—to gain purchase in a rationally-ordered universe, there must be something for the will to *do* in ordinary human choice-making, a possibility that seems to be ruled out by both the ancient and early modern rationalist dogmas. If the response of the will to normative pressure is akin to the response of ordinary objects to physical pressure, its operation is essentially *hydraulic*: it contributes

[38] Al-Ghazâlî's work is quoted *in extenso* in Averroes's century commentary thereon, *The Incoherence of the Incoherence* (*Tahâfut al-Tahâfut*). Averroes 1954. Al-Ghazâlî uses the example of a man facing a choice between two (virtually) identical pieces of fruit as a way of motivating his claims about God's freedom in willing (Averroes 1954, vol. I, pp. 20–25). His full account of the power of human willing holds that it in fact lacks this freedom; this view is explicated more fully in his *Book of Faith in Divine Unity and Trust in Divine Providence* than it is in *The Incoherence of the Philosophers*. For a detailed account of Al-Ghazâlî's view of human freedom in the former, see Burrell 1999; see also Kaye 2004, esp. p. 22.

nothing to choice itself except, when invoked at all, to serve as a passive conduit through which normative pressures are translated into action.[39]

2.2 Dogmatic rationalism

Classical rationalism, as we have defined it thus far, involves commitment to just two theses about rational choice. The first is the idea that rational standards are essentially *comparative*—a choice is rational just in case it involves selection the *most* choiceworthy option, where an option's choiceworthiness is understood as its being univocally favored by the balance of reasons at hand. The second is the conflation of these rational standards with the standards for paradigmatically intelligible human action. There is something incomprehensible about choices that take as their object something other than the best. The balance of reasons constitutes choices as intelligible and marks them as rational at the same time.

Any number of further theses about the world and our place in it might be paired with the classical perspective on choice. There are therefore manifold ways to be a classicist about rational choice. Nothing about classical rationalism, as defined, entails that the world needs to be a certain way. Perhaps the nature of human self-consciousness demands a classical perspective on choices made, but reality, in all its messiness and beautiful

[39] Use of the term 'hydraulic' to describe certain models of human motivational psychology and will is not original to me, of course. I was reminded near the completion of this project of where I first encountered the description: when it was used by R. Jay Wallace to describe internalist (and meta-internalist) moral psychology, according to which the connection between *desires* and action is as exceptionless and basically automatic as is the connection between *reasons* or *rational judgments* and action is for the dogmatic rationalists. See Wallace 1999, p. 234, esp. n. 24, where he insists that hydraulicism is guilty of "…tracing our actions and inferences to the causal operations within us of psychological forces with respect to which we as agents are essentially passive."

complexity cannot answer that demand. Perhaps the demands of practical reasoning themselves structure the world as it presents itself to us for deliberation and choice, so that constraints on the intelligibility of choice themselves constrain, in a deep way, reality as we can possibly inhabit it. Each of these possibilities poses a problem for the classical model of rational choice, of course, which we will explore in due course. For now, though, we will turn our attention to one particular way of being a classical rationalist: the one on display in the work of the early modern rationalists already discussed, which we might label 'dogmatic rationalism.'

Dogmatic rationalism pairs the classical model with a picture of the world within which human movement takes place which *explains why* the two theses about rationality and agency are true—namely, by subsuming them as special cases of more general rules governing the causal or rational structure of the universe.[40] This pairing, as we have seen via our dogmatic exemplars, Leibniz and Spinoza, has the virtue of situating our explanation of human choice within a systematic and coherent framework, and of offering at least some justification for the classical model. But, as we shall see, it also robs the dogmatist of the tools for explaining how deviations from the classical model's predictions are so much as possible. Since the world is itself structured by the Principle of the Best, it will not present agents with underdetermined choice contexts; if, through some error or cognitive limitation, we *perceive* a choice as underdetermined,[41] we should (and will) be immobilized by this perception. In the absence of a compelling reason to move forward

[40] McDonough 2011, p. 26. (Page reference is to manuscript version).

[41] As is arguably the case in Leibniz's own discussion of sin and weakness of will—see below.

one way rather than another—or in the absence of some tiny perception of a difference in choiceworthiness of options, even the merest *twinge*[42]—there can be no movement.

As we shall see, there are problems both with the dogmatic rationalism of the early modern rationalists, and (in §3) with the hybrid voluntarism of many medieval theologians and early modern moralists in between. This model of willing not only flouts common experience, it poses serious problems for the possibility of free agency. Eventually, we will find that even a refined and updated version of the view is subject to such concerns.

To begin, consider Spinoza's take on the problem of indifference:

> I entirely grant that if a man were placed in such a state of equilibrium he would perish of hunger and thirst, supposing he perceived nothing but hunger and thirst, and the food and drink were equidistant from him. If you ask me whether such a man would be thought an ass rather than a man, I reply that I do not know; nor do I know what ought to be thought of a man who hangs himself, or of children, fools, and madmen.[43]

Spinoza straightforwardly admits that his system lacks the resources to account for deviations from perfect rationality. What is there to be said of *akrasia*, of the development of rationality from birth to maturity, of simple mistakes in ordinary reasoning? Not much, it seems—on Spinoza's hermetically-sealed panentheism, all of nature is ordered from top to bottom by a divinely-organized rational determinism. Whatever is, is in God;[44] all

---

[42] Leibniz's technical term for such 'twinges' is *petites* perceptions, perceptions of which we are not conscious at the time of occurrence (*New Essays,* A iv. VI, 45; see also A vi. VI, 118). These can be causally efficacious in moving us to act, though they do not involve the conscious perception of a reason for so acting, and so they are crucial to explaining some choices. More on this below.

[43] *Ethics*, IIp49.

[44] *Ethics*, Ip15: "Whatever exists exists in God, and nothing can exist or be conceived without God."

existent things and events are inevitable modal emanations of God; and every outcome, from the tiniest human choice to the grandest cosmic eruption, proceeds with absolute necessity.[45] Like the ancient Stoics, Spinoza derives his ethics from his metaphysics, and like them has no little difficulty accounting for deviations from perfect virtue and rationality, given that we are all emanations from a perfectly rational world spirit or *Logos*.[46]

Leibniz is in many ways no less dogmatic, but in the case of human choice he does at least attempt to offer an explanation for cases of apparent tie-breaking in cases like **buridan's ass** rather than simply admitting the failure of his system to so much as explain the possibility of such phenomena. He insists on the necessity of inactivity when reasons underdetermine choice, as we see in his correspondence with Samuel Clarke:

> In things which are absolutely indifferent there can be no choice and consequently no election or will, since choice must have some reason or principle.[47]

Or again:

> To say that the mind may have good reasons for acting when it has no motives, and when things are absolutely indifferent … is a manifest contradiction. For if there are good reasons for the course it adopts, the things are not indifferent to it.[48]

---

[45] *Ethics,* Ip29: "In nature there is nothing contingent, but all things have been determined from the necessity of the divine nature to exist and produce an effect in a certain way." *Cf.* Ip33.

[46] There are many other fascinating similarities between Spinoza's panentheistic rationalism and the pantheistic rationalism of the ancient Stoics. For discussion see James 1993; DeBrabander 2007.

[47] Leibniz and Clarke, 1715, §1 of the fourth letter.

[48] Leibniz and Clarke, 1715, §16 of the fifth letter.

Leibniz takes the fact that choices sometimes seem to be made under ostensibly indifferent conditions to be evidence that such choices were not really underdetermined after all, but was rather driven by imperceptible distinctions between the apparently normatively identical options, which serve to render a normative distinction between them.[49] If things had been genuinely underdetermined, he reasons, no action would have been taken at all. Just as in the divine case, the evidence of choice is sufficient to show that can have been no strict equivalence of alternatives in the first place. The justification Leibniz offers is this:

> There is never any *indifference of equipoise*, that is [situations of choice] where all is completely even on both sides, without any inclination towards either. … [T]he case also of Buridan's ass between two meadows, impelled equally towards each of them, is a fiction that cannot occur in the universe, in the order of Nature… For the universe cannot be halved by a plane through the middle of the ass, which is cut vertically through its length, so that all is equal and alike on both sides. […] Neither the parts of the universe nor the viscera of the animal are alike, nor are they evenly placed on both sides of this vertical plane. There will therefore always be many things in the ass and outside the ass, although they may not be apparent to us, which will determine him to go to one side rather than the other. And although man is free, and the ass is not, nevertheless for the same reason it must be

[49] *Cf.* Leibniz *New Essays*, Bk. II ch. xxi §48: "*Th.* [To speak accurately, one is never indifferent regarding two alternatives, whatever they may propose… for we do the one or the other without thinking of it, and this is an indication that a concurrence of internal dispositions and external impressions (although insensible) determines us to the side that we take. But the prevalence is very small, and in case of need it is as if we were indifferent in this respect, since the last sensible subject which presents itself to us is capable of determining us without difficulty to one rather than to the other…]" (pp. 204–205). Theophilus, speaking here, represents throughout the rationalist point of view, against the empiricism of his interlocutor, Philalethes.

> true that in man likewise the case of a perfect equipoise between two courses is impossible.[50]

Like the Leibnizian solution to the divine paradox I've named for him (**leibniz's god**), the philosopher's conclusion is again simply to deny the problem. Underdetermination of *any* choice—divine, human, or animal—is ruled out by his metaphysics, and so the only move available is to explain the phenomenon away. There is simply no room to accommodate it.

And so extraordinarily counterintuitive things are countenanced on the dogmatic view, indeed, actively embraced because of their alleged logical inevitability.[51] The rationalists offer a twofold rejection of the possibility underdetermined choice. First, according to Leibniz at least, there can be no choice contexts which manifest this structure, since there must always be a determinate normative fact of the matter about what is best. Second, if one were, *per impossibile*, to face such a choice, one indeed rationally ought to be paralyzed by indecision. God would do the same, and in fact *could not act* in the absence of some fact of the matter about what is best. It is part of the nature of choice that it could not proceed in underdetermined contexts: it must have some reason or principle which is sufficient to explain the election of one course of action over another—whether introspectively

---

[50] Leibniz, *Theodicy*, Bk. II ch. xxi §§46–49. The same notion reappears in his correspondence with Pierre Coste: "The universe is incapable of being divided or split up into two equal and similar parts. The universe is not like an ellipse or other such oval, where a straight line drawn through its center can cut it into two congruent parts. The universe has no center, and its parts are infinitely varied; thus the case never arises in which everything is perfectly equal and strikes equally on all sides" (Correspondence with Coste, "On Human Freedom," in *Philosophical Essays,* as quoted in Jauernig 2008, p. 201).

[51] From Michael Della Rocca's charming description of his favored dogmatist: "…*Spinoza* is not one to avoid biting bullets or, more accurately, what one might see as bullets the biting of which is to be avoided, Spinoza often sees as logical or rational conclusions to be embraced because of their rationality, because of their logical inevitability." Della Rocca 2008, emphasis in original.

available at the moment of choice or no.[52] This is true no less for divine than it is for human decision-making. Since God seems to have resolved what we might have thought to have been some serious **indeterminacy**, his choice must not have been underdetermined after all. Since humans and animals often act when options seem to be **indifferent**, they must not really be so.[53]

Fools, madmen, and children walk in our midst—but what can be said of their inner life, really? They are anomalies rather than data to be accommodated by the beautifully self-contained dogmatic worldview. The identification of strong comparativist standards for rational choice with those for coherent action flows naturally from the dogmatic insistence that the causal structure of the world itself be understood in terms of conceivability, from their subsumption of the explanation of human activity under laws which govern the entire universe.[54] When rational intelligibility is a standard to which any fact or event must measure up, action performed in defiance of the balance of reasons is, quite literally, inexplicable except in terms which reduce it to something other than self-conscious agency. The world itself cannot but live up to the Principle of the Best; when our own choices fail to do so—as they so often must—their unintelligibility is a violation of the natural order, which can countenance neither irrationality nor arationality. Spinoza wonders:

---

[52] Leibniz and Clarke, 1715, §1 of the fourth letter.

[53] And in both cases, for Leibniz at least, the normative differentiation of apparently undifferentiated options is tied to his doctrine of the identity of indiscernibles (II), though it is not clear whether the relationship between PSR and II is one of mutual entailment or if he instead uses the one to explain the other. For discussion see Mates 1986; Melamed and Lin 2011; Jauernig 2008; McDonough 2011.

[54] *Cf.* Garrett 2003, Melamed 2012, for challenges (or at least a contrast) to this view.

> [W]hat altar of refuge can a man find for himself when he commits treason against the majesty of reason?[55]

What refuge, indeed? There is nowhere to go. For the dogmatists, the majesty of reason is inviolate, exhaustive, and ubiquitous. Their view ascribes to reason the same organizing and governing role that traditional theologies ascribe to God.[56] Nothing exceeds its strictly comparativist reach, so there is no room for contingency in the normative structure of even the most unimportant choices.

2.3 Problems for the view

Our quick tour of classical and early modern rationalism has necessarily glossed over significant distinctions between and within the views, sacrificing depth and nuance for concision and thematic unity. Of interest to us is the way that the conflation of rational and constitutive standards for human choice engendered serious problems for some of the historically prominent models which made it a centerpiece. But the dogmatism of the early modern period (and probably that of the Stoics as well) is not a consequence of the classical model—it flows from what stands behind it and gives it life in the early modern context.

The rejection of the possibility of underdetermination is the source of some of the most serious challenges to dogmatic rationalism. First, as noted above, it is manifestly implausible to insist that rationally underdetermined choice contexts are not so much as a theoretical possibility, as both Leibniz and Spinoza do. It *may* be true, as Leibniz argues,

---

[55] Spinoza, *Theological-Political Treatise,* III/188.

[56] Melamed 2010, pp. 129–130.

that we are never presented with truly qualitatively identical options, as is presupposed by one description of the problem of **buridan's ass** (and is, indeed, entailed by another of his key precepts, the Identity of Indiscernibles[57]). It may be true, that is, that there is likely to be *some* distinction (besides spatial differentiation) either in the bales of hay or the viscera of the animal that rules out literal indistinguishability. Such explanations will surely be *less* ready-to-hand, however, for many choices made possible in the modern world: a forced choice between visually indistinguishable HEX codes for a website element, or from a row of factory-machined drawer-pulls. And even if we could still find *some* criterion to differentiate them, the distinctions which mark options as non-identical might nonetheless fail to present themselves to us as perceptible, beneath a certain threshold—and therefore, it would seem, fail to enter the normative calculus at all.[58]

Indeed, even supposing that the claim about **buridan's ass** holds, it is no argument against the possibility of a situation like that of **aristotle's dog**, where what is problematic is not an equivalence of options but an equivalence of *choiceworthiness*.[59] Against such a possibility, one imagines Leibniz again doubling down on the claim of differentiability, inferring from the fact that a choice is made that there was some distinction between options, however

---

[57] Leibniz, *New Essays,* A iv. VI, 42. Spinoza, too, was a proponent of the identity of indiscernibles: see, e.g., *Ethics* EIp4.

[58] One might also suppose that these distinctions, even if perceptible, will be so insignificant as to fail to contribute to the normative calculus: considerations which count as reasons must, on this view, earn their right to the name. Actually, as noted in §**I**, I think views of this kind are mistaken —but if one were to hold this view, it would provide one more route to rejecting the dogmatic denial of underdetermination.

[59] **Incomparability** and **incommensurability** are not live theoretical possibilities for the dogmatist, for obvious reasons—when all of the world is subject to precisely the same sort of conceivability claims, the comparability of alternatives is not a problem, for the material of comparison applies equally to all.

imperceptible our apprehension of that distinction was at the moment of choice. Such a view is suggested by some of his comments about *petites* perceptions, mental events which can be causally efficacious in moving us to act although they do not involve the concurrent (conscious) perception of a reason for so acting:

> All our unpremeditated actions are the result of a concurrence of *petites perceptions*, and even our habits and our passions, which so much influence our [conscious] deliberations, come therefrom… I see that among those who discuss freedom of the will there are some who, taking no notice of these unperceived impressions which are capable of inclining the balance, imagine an entire indifference in moral actions, like that of the ass of Buridan equally torn between two meadows.[60]

With this conceptual machinery in place, we see that the doctrine rejecting genuine underdetermination has become unverifiable, and therefore (at the very least) explanatorily inert. Any instance in which an agent seems to break or ignore a normative tie turns out to be one which can be explained, retrospectively, in terms of some distinction in options whose existence we infer even though it was not consciously available at the time of choice.[61] This is so even in those cases where supposing that one option of the choice set was genuinely more choiceworthy than others is extremely difficult—as is the case with the

---

[60] Leibniz, *New Essays*, II, i, 15. Reliance on *petites* perceptions is apparently what enables him to insist that "There is indifference, when there is no more reason for one than for the other. The opposite is determination… All actions are determined, and never indifferent. For there is always a reason inclining us to one rather than the other, since nothing happens without a reason… a liberty of indifference is impossible. So it cannot be found anywhere, not even in God. For God is self-determined to do always the best. And creatures are always determined by internal or external reasons." (*Philosophische Schriften*, vol. 7, p. 109), as first quoted in Russell 1900, and then quoted by by Rescher 1960 fn. 78, pp. 29–30.) Here, by 'internal reasons' we might presume that he refers to these very same *petites* perceptions.

[61] For an argument for why *petites* perceptions must be introspectively available after a certain interval, see Kulstad 1977.

rational paradox of **leibniz's god**, to which, as we saw, Leibniz also offered only a dogmatic reply.

And so we reach the second half of the twofold rejection: denial of the claim that forward movement is so much as possible when one is unable to scry a meaningful distinction in the choiceworthiness of options. When we fail to deliberate, when we act without thinking, *petites* perceptions step in as a kind of conceptual proxy for our apprehension of reasons for so acting, and make possible (and indeed seem on Leibniz's view to justify) our movement. In those cases where we deliberate, feel genuinely torn, and then choose anyway, there must have been some inclination of which we were not aware at the time that explains how we were able to move one way rather than another. And when the underdetermination presented by the choice set at hand looks all but insuperable—as in the infinite series of identical moments in time from which **leibniz's god** might be expected to have selected for the world's creation—we must simply take it on faith that there was some difference there.

The first half of the twofold rejection is a claim about the way the world is organized: a claim at least supported, and more probably entailed, by the nexus of metaphysical claims which comprise dogmatic rationalism's conceptual substructure. The second half is a claim about the rational will itself, rather than the world in which it operates. Can a rational being move forward from deliberation to choice when options are, from her perspective, incomparable, vaguely-specified, equally-favored, and so on? The dogmatists are apparently committed to denying this possibility. The will's activity consists in responding to the balance of reasons, and translating it into action; in the absence of a clear balance, there

can be no forward motion. It is the second half of the twofold rejection of underdetermination which troubled many of those otherwise fully committed to the classical model of rational choice—who took it upon themselves to articulate what was problematic about this model of willing, and to defend a version of classical rationalism on which the will's contribution to the process of human choice-making was something other than merely hydraulic.

What I am calling 'hydraulicism' about the will—an understanding of its role in rational choice-making as essentially passive, serving as a kind of conduit along which rational pressures are translated into action—is an animating feature of the dogmatic rationalism of the early modern period. As already noted, it is hard to countenance the possibility of deviations from perfect rationality in human choice-making on such a view; to take just one example, the phenomenology of *akrasia* or weakness of will is not easily assimilated to theoretical error, which would seem to be the only route open to the dogmatic rationalist. Hydraulicism clashes with our sense of how our choices proceed: we seem perfectly able to acknowledge where the weight of reasons lies in rationally-determinate choice contexts, and yet act otherwise, whether out of weakness or sheer willfulness.[62] And while hydraulicism was one of only many aspects of the dogmatic view that clashed with common sense, it was a crucial sticking point for many because it *also* clashed with conceptual desiderata for human choice—that it be 'self-directed,' or 'owned,' or 'authored,' or 'free'—which flowed from other bedrock theoretical commitments.

---

[62] As an enormous literature in contemporary action theory manifestly demonstrates—this is a strong sense that people have about how they make choices, and just what weakness of will (etc.) involves: Buss 1997 is both an exemplar and a useful overview of related ideas.

It was this need that drove the modified rationalism of medieval scholastics and early modern moralists, who embraced the guise of the good alongside other commitments that required a more robust conception of free agency than the dogmatic view could countenance.[63] Christian theologians, concerned to accommodate theistic tenets like sin and the possibility of divine grace, required a theory of human choice with some room to explain human failure, and the value of overcoming it. They required the space to turn away from God and the space to turn back, as well as a way of explaining the goodness of one and the evil of the other. Early modern British moralists responsible for what Stephen Darwall has termed 'the internalist turn' in ethics were driven instead by a desire to account for the directed and personal nature of moral responsibility.[64] They, too, sought to explain the value of morally good choices in terms of the *freedom* to choose otherwise—a freedom which the will is denied on a hydraulic model of its operation.

For the moment, let us set aside the phenomenological challenge to hydraulicism, to focus on the notion that a purely hydraulic model of the will's operation rules out the kind of freedom in choice-making that would seem to be required for the *ownership* or *authorship* of human action required for personal (moral, theistic) responsibility. Spinoza, as we have seen, was perfectly happy to reject the possibility of free will: not even God has liberty in choice. He is a hard determinist through and through. We saw, too, how al-Ghazâlî endorsed a kind of libertarian freedom of the will for God, but denied it to humans.

---

[63] And what of the classical rationalists who opened the chapter for us: Socrates, Plato, Aristotle? Not exactly dogmatists, they are not in any meaningful sense part of the voluntarist movement canvassed here and in the next two chapters, mostly because the moral psychology implicit in their disquisitions lacks a notion of the 'will' as an active and explanatorily unique principle in choice Dilhe 1982. More on this in §**3**.

[64] Darwall 1995.

While Leibniz rejects the possibility that God could make an arbitrary choice (or make an arbitrary choice into a reasoned one through some exercise of will), he nevertheless makes it a point to insist, contra Spinoza, that God's will is free. Though God cannot act in any way other than the best, and though the world will never present God with any choice that does not have a single best outcome, this is compatible with free choice. Indeed, acting for the best is the most perfect manifestation of God's freedom:

> *Th.* It seems to me that, properly speaking, although volitions are contingent, *necessity* should not be opposed to volition, but to *contingency* … and that necessity should not be confounded with determination […] Geometrical and metaphysical consequences necessitate, but physical and moral incline without necessitating; the physical even having something of the moral and voluntary as related to God, since the laws of movement have no other necessity than that of the best. Now God chooses freely although he is determined to choose the best…[65]

This conception of freedom extends to human choice as well, where Leibniz again makes the familiar compatibilist point that necessity in choice—even necessary determination by a normative balance outside of the self—should be contrasted not with freedom but with arbitrariness. The only freedom of the will worth wanting is the freedom to "act according to the choice of [our] own mind," and anyone who seeks another freedom "knows not what they ask."[66]

---

[65] Leibniz, *New Essays*, Bk. II, ch. xxi, §13, pp. 182–183.

[66] Leibniz, *New Essays*, Bk. II ch. xxi §8, p. 179; §48, p. 205. For an extended discussion of these themes, see further Bk. II ch. xxi §§5–8, 12–17, 19, 47–50.

Much of what Leibniz has to say about freedom of the will, including his extended discussion of weakness of will, suggests a kind of voluntarism.[67] This is especially true of his most famous claim about the way in which reasons or motives operate upon the will: that they "incline without necessitating," as in the passage above.[68] Elsewhere:

> For speaking absolutely, our will is in a state of indifference, in so far as indifference is opposed to necessity, and it has the power to do otherwise, or to suspend its action altogether, both alternatives being and remaining possible.[69]

If the balance of reasons inclines but does not compel choice, if the connection between judgments of most reason and action is neither causal nor metaphysical, then how does choice proceed? It would seem to require the input of some third factor—the impetus of the will, which directs choice—for any particular choice to go forward. And what is the will's contribution? In order for there to be some contribution at all, it would seem, the will must be able to genuinely determine the course taken, whether in accordance with or in defiance of the balance of reasons. But Leibniz denies this:

> …but choice, however determined the will be, should not be called necessarily and rigorously absolute; the prevalence of perceived good inclines without

---

[67] For extended discussion, see Vailati 1990, where Leibniz's epistemological account of the possibility of weakness of the will is unpacked and meant to explain the existence, prevalence, and phenomenology of *akrasia* without abandoning commitment to the Principle of the Best.

[68] The claim is reiterated in many places, including, *inter alia*, Leibniz, *New Essays,* Bk. II ch. xxi §§8, 12, 13, 49.

[69] Leibniz, *Discourse on Metaphysics*, §30. And yet just a few lines later, he insists "It is true, however, and indeed it is certain from all eternity, that a particular soul will not make use of this power on such and such an occasion." See also his remark in "Necessary and Contingent Truths" that free or intelligent substances like humans and God "are not bound by any certain subordinate laws of the universe, but act as it were by a private miracle, on the sole initiative of their own power" (p. 100).

> necessitating, although considered as a whole, this inclination is determinate and never fails to produce its effect.[70]

It is difficult to see in what sense this could constitute the kind of freedom of will sought by most who deem it important. Where the moment of decision arrives, "everything which impinges on us weighs in the balance and contributes to determining a resultant direction, almost as in mechanics."[71] This is why, as we saw in the preceding section, absolute paralysis would necessarily occur were the world to present a rational actor with a truly underdetermined choice context. There simply are no resources outside of the rational pressure of the balance of reasons to drive choice. We can sometimes be mistaken about what is in fact best, and we can also fail to appreciate vividly that which we 'know,' in some sense, to be best, so that our motivations become confused. For Leibniz, unlike for Spinoza, both mistakes and weakness of the will are possible. But this is not because the will itself steps in to override the deliverances of the intellect: it is only due to some mistake in the intellect's apprehension of the normative contours of the case in view.[72] The will does nothing but respond in the way it normally does to these deliverances, namely, complete capitulation. So although there is something called 'will' on Leibniz's picture, it adds nothing to the mechanism of choice. It has been written right out of the model: there is nothing left for the will to *do* or *be* except a conduit for the translation of reasons into choice, and so talk of its freedom is, as Leibniz himself sometimes seems to admit, more than a little misleading.

---

[70] Leibniz. *New Essays*, Bk. II ch. xxi §49, p. 206.

[71] Leibniz, *New Essays*, Bk. II ch. xxi §47, p. 202.

[72] DeBrabander 2007.

2.4 The missing agent

It is worth pausing for a moment to consider the presuppositions about what matters in human choice which clearly underwrite both classical rationalism generally and dogmatic rationalism in particular. The threat of underdetermination is that it leaves us with arbitrariness. Inescapably arbitrary choice means that there can be no comparative basis for the choice made and therefore no coherent story about why we moved the way that we did in the world. If arbitrariness in any thread of the fabric of the world is impossible, then it must also be impossible in human choice, and appearances to the contrary—whether based on the first-personal phenomenology of willful or *akratic* choice, or the third-personal observation of foolishness and sin in the world around us—must be explained away. And it is worth accepting that one's model of human choice flies in the face of common experience *and* important conceptual desiderata about freedom and self-determination, because the theory as a whole will have preserved that which is most essential to our understanding of human agency: intelligibility and coherence in choice, which can only come from strong comparativist rational determination.

The objection to dogmatic rationalism we've been circling around is the notion that our model of rational choice is somehow missing the character one might have thought to be the linchpin of the whole: the agent who makes the choice. The complaint can be stated in at least two ways, each of which objects directly to the hydraulicism of the dogmatic picture of willing. On the one hand, the rational–causal[73] determinism which, for the dogmatists, was a key driver of the identification of rational with constitutive standards for

[73] Recall the dogmatic identification of reasons with causes, and much else besides.

human choice, cannot allow that human actions are at least sometimes, or at least in part, self-caused. They rule out in a very direct way the possibility of an agential contribution to action. The picture of human agents, on the deterministic model, is as a kind of blank space for the interplay of causal forces:

freewill$_1$	Causal determinism rules out the possibility of self-caused action, and this is a threat to the possibility of free agency.

The Leibnizian response to this kind of objection is perhaps sufficient. Seeking freedom in randomness, looking for the possibility of agent causation in quantum indeterminacies, as some contemporary free-will libertarians have done, seems to be a kind of category mistake. There have been numerous, highly sophisticated attempts to explain how libertarians can wring meaning from causal indeterminism, including those which make explicit reference to the importance of reasons-explanations in securing this meaning.[74]

But the causal–rational determinism of the dogmatic rationalist also rules out another, subtler dimension of human agency. The drive to understand our actions as free is surely in some sense the desire to see our wills as causally efficacious, in a way that is prohibited, on first glance, by causal determinism. But often we are also driven to see ourselves as acting freely because we want to see our actions as *our own*, as expressing our own desires, inclinations, and ideals. This is precisely the concern that drove the morally- and theologically-motivated theorists mentioned above. The evil of sin, the personal nature of moral attributability, the need for divine grace—each seems to require an ownership of our actions and choices that is ruled out by the hydraulic model. To understand what goes

[74] The efforts of Robert Kane (1985, 1996) are notable on this front in particular; for a useful survey of other contemporary work see Clarke 2000.

wrong when we deviate from our own judgments of what is good or what is best, we must at least sometimes be able to see these deviations as driven by something other than simple misapprehension of the good.

And so, very loosely stated, the deeper concern about dogmatic rationalism is this:

freewill$_2$	Rational determinism rules out the possibility of self-governed action, and this is a threat to the possibility of moral agency.

We will unpack this objection in the next chapter, for it was the driving concern of many of the classical rationalists who became free-will voluntarists rather than dogmatists. One thing to note for now, however, is that the 'rational determinism' to which the concern refers has nothing to do with causal determinism. The metaphysics of dogmatic rationalism meant that causal and rational determinism amounted to about the same thing —but here the only element of concern is the thesis about the nature of human choice itself. Let all the metaphysical and theistic machinery fall away: if human decision-making proceeds in the merely hydraulic way that is suggested by the identification of constitutive with rational standards for choice, the problem remains. Agents are nothing more than loci for the interplay of rational forces, an interplay governed by laws not of the agent's making.

2.5 Conclusion

Certainly classical rationalism does not entail its dogmatic reduction. The problem we have uncovered is not that classical rationalism fails as an account of how human choice often proceeds, when it is going as it should; the failure belongs instead to the structures

used to explain or justify it. When these entail claims that clash with common sense, they ought to be reformed or rejected outright. But surely there is something *right* about a model of rational choice on which irrational action is essentially unintelligible; the error of the dogmatists was in thinking that intelligibility is a demand we can make of the universe and not just of ourselves.

Can the classical model accommodate a non-hydraulic conception of the will? This question drives the investigation of the next two chapters, which look at the effort to reintroduce a notion of the 'will' as an effective cause in human action to a model of rational choice which does not naturally accommodate it. For the dogmatist, the intelligibility-conferring power of rational justification is all we have reason to want in our model of human agency. To look for more in an account of the way humans make their way in the world is to "revert to the chimerical, as in the empty faculties or occult qualities of the scholastics, in which there is neither rhyme nor reason."[75] But this reversion would seem to be exactly what is called for to resolve the problem of the missing agent. Where there is no will, the thought goes, there is no agency; the effort to include the will is an effort to rediscover an agent who has apparently gone missing within a nexus of irresistible causal and rational forces.

The voluntarist response to dogmatic rationalism and the hydraulic model of the will insists that the will *can* and *does* contribute to the outcome of choice, in a way that

[75] Leibniz, *New Essays*, II, xxi, p. 203: "We [can] become as it were masters of ourselves, and make ourselves think and do at the time as we should wish to will and as reason commands. But it is always through determined paths, and never without a reason, or by means of the imaginary principle of perfect indifference or equilibrium … To assert otherwise is to revert to the chimerical, as in the empty faculties or occult qualities of the scholastics, in which there is neither rhyme nor reason."

(whatever its causal–metaphysical status) is meant to secure the appropriate relationship between agent and outcome. This voluntarism proceeds, oddly enough, by revisiting our old friends the **dog** and the **ass**—the primary models of underdetermination countenanced on the classical model. The voluntarist suggestion is not just that rational underdetermination of this kind is possible, but that it alone invites the kind of volitional contribution to choice that is demanded by **freewill**$_2$. When the balance of reasons is determinate, we are indeed driven to follow its dictates in much the same way envisioned by the dogmatists; but when reasons run out, our will can somehow bridge the gap between deliberation and action. These underdetermined choices give us the chance to express and perhaps even forge our individual characters, moral or otherwise, in a way that renders us the *authors* of our actions. It is only this self-understanding as creators rather than mere actors that makes sense of a variety of central moral and theistic tenets—and, perhaps, of our own more nebulously-stated desire to discover *the agent* somewhere in our account of how choice proceeds.

3 Hybrid voluntarism

Albrecht Dihle popularized the notion that the concept of 'will' as a distinct human faculty did not exist in classical antiquity. While *akrasia* and willful wrongdoing were recognized phenomena in the Greek tradition, they were typically described as failures of reason, rather than failures of will. Good action depended on good reasoning, and any 'change of heart' was viewed unfavorably only because "it always implie[d] that one of the two intentions is or was based in error."[76] The moral psychology of the classical world had no space for a spark of deliberate self-movement; agents were impelled from without by

[76] Dilhe 1982, pp. 30–31. Compare this linguistic note: "Our term 'will' … leaves out any special reference to thought, instinct, or emotion as the cause of that intention. Greek, on the other hand, is able to express intention only together with one of its causes, but never in its own right" (pp. 24–25).

the perceived goodness of good things or driven from within by irrational passion—a splinter of the self rather than the whole of one's being.[77]

It was the notions of obedience and forgiveness central to the Abrahamic faiths that first called for a faculty of 'will' as an essential part of the human constitution. Indeed, the tradition made will, rather than reason, central to its very cosmology: God created the world and ordered it from top to bottom through self-directed choice, not reason alone.[78] Though the divine will is perfectly rational, it is primarily that will, and not God's reason, that must be obeyed, as the parable of Abraham and Isaac well illustrates.[79] The "beginning of wisdom" is a capitulation to God's volition, not an apprehension of the cosmic order God has created.[80] Since only God's will is apprehensible by humankind, and

---

[77] Aristotle is a complicated case, though I have lumped him in here with the rest of the classical tradition: he differentiates, more than his teachers and more than later classical thinkers such as the Stoics and Epicureans, between the faculties of theoretical and practical reason, and has more nuanced and complicated things to say about the possibility of distinctively practical judgment (and practical error). But even he does not have a theory of the will as such, nor does his model leave space for one. See Dilhe 1982, pp. 55–61; *cf.* Moss 2010.

[78] Dilhe 1982, pp. 15–17.

[79] Genesis 22:1–18.

[80] The idea that "fear of the Lord is the beginning of wisdom" is woven through the Old Testament, see, e.g., Psalms 111:10; Job 28:28; Proverbs 1:7.

not God's reasons,[81] the human response to divine will is itself another exercise of will—neither apprehension, perception, nor cognition, but simple capitulation.[82]

As the Christian tradition blossomed from and alongside the Old Testament faith, notions of grace and forgiveness became central to the common understanding of humanity's place in the world. St. Paul's description of deliberate wrongdoing is instructive: sin is a failure of will from which no amount of knowledge can save us, even knowledge of the will of God. The doctrine of original sin dooms us to deviant intentions. Only the gift of divine grace can save us from our sinful inclinations, and we display the virtue of humility in acceptance of our fallen nature and gratitude for divine help.[83] Yet since our will determines our action, it is subject to divine moral judgment all the same—willed action is attributable to its author as a whole in a way that action based on cognitive error or loss of control is not.[84] Later Christian thinkers like St. Augustine of Hippo and St. Thomas Aquinas made this notion of will central to their understanding of the relationship between God and individual,[85] as well as to moral dimensions of the human experience like the sin, grace, and redemption.

---

[81] "For your thoughts are not my thoughts, saith the Lord." Isaiah 55:8.

[82] Dilhe 1982, pp. 75–77. Compare Seneca's remark, representative of the Stoic insistence that compliance with the divine is a matter of reasoned action on the basis of calm perception of the rational order of the world: *Non pareo deo sed assentior*—roughly: I do not obey God, rather, I agree or comply (Seneca, *Epistulae Morales* 96.2, author's translation*). There is no need for blind obedience in the Stoic view, for the divine instructions on how to live (if we may call them that) are woven into the fabric of the cosmos itself and accessible to anyone who takes the time and care to ascertain them. For more, see Brennan 2005. [**technically, author's sister's translation.*]

[83] Romans 7:7*ff*. For extended discussion, see Dilhe 1982, pp. 70–81.

[84] Dilhe 1982, pp. 83–90, esp. pp. 84–85.

[85] According to Dilhe, Augustine invented the concept of will (*voluntas*) as we know it today. Dilhe 1982, p. 123. For further discussion, see pp. 125–144.

If accurate, this tale of the birth of the will in Western philosophy would provide a tidy explanation for the unconcern of classical thinkers with certain issues about intentionality in agency, as well as for the state in which we found the dogmatic rationalists: offering a picture of rational choice and human agency in which the will served as an empty concept, a vestigial organ of Christianized moral psychology for which there is no real use in the dogmatic reclamation of robust classical rationalism. At any rate, some idea of volition was conceptually required by the moral systems of two important contrast classes to the rationalists already considered: medieval Christian scholastics and early modern British moralists.

3.1 The voluntarist turn

The voluntarist, as I am defining her, is a rationalist at heart who finds herself dissatisfied with the hydraulicism of the classical model of choice. Like the classicist, she hews to a comparativist standard of rational choice, and like the classicist, she sees the balance of reasons as intelligibility-conferring and a marker of specially human choice. But unlike many classical rationalists—certainly unlike the dogmatists—she identifies *underdetermined* choice-making as an important exercise of human agency. Whether underdetermined choice-making constitutes the whole of human agency or just a part, or whether this part is more authentically an expression of our 'selves' than rationally-determinate choices (or vice versa), are matters on which various tribes of voluntarists disagree. Just as there are many ways to be a classical rationalist, there are many ways to be a voluntarist.

The motivations of those swept up in the voluntarist turn, however, exhibit a strong family resemblance. Medieval theologians and early modern moralists concerned with freedom of choice were troubled by the model of willing suggested by classical theses like the guise of the good and the principle of sufficient reason. They often shared certain ideas about the world with the dogmatists—its Providential ordering, for instance—but sought to find a conception of the will's contribution to choice that could explain the possibility (not to mention the prevalence) of sin, the wickedness of immoral choice, and the specifically moral value of actively pursuing a good to which we do not find ourselves automatically inclined. The account of weakness of will available on the dogmatic view is as unsatisfactory as the Socratic explanation of vice: in both cases what we might have thought was a failure of moral or rational agency is simply reduced to a theoretical mistake. For the voluntarists, this simply will not do.

Our focus in this chapter is on a tribe of voluntarists I will call *hybrid voluntarists.* Like any voluntarist, the hybrid voluntarist seeks to square the intuitive plausibility of the comparativist standard with the demand that our agency consist in more than simply (and unerringly) translating rational directives into action. The hybrid solution is to tack onto the hydraulicism of the dogmatic model a capacity for willed choice in those cases where reason underdetermines the normatively appropriate outcome.

This is a fine needle to thread, for a number of reasons. One we have already seen: the tighter one binds the constitutive standards for human choice itself with the merely evaluative standards for rational choice, the more difficult it is to find space for recognizably human deviations. Yet violations of the standards for rational choice seem to

involve, by very definition, a departure from ordinary sources of intelligibility and self-understanding in choice, *viz.*, the reasons for which we act, and their combined weight relative to other available options. In order to secure the sense of responsibility required for moral and theistic purposes, the voluntarist must take care to articulate a notion of freedom worth wanting—one which does not locate the supposedly-valuable activity of the will at precisely that place where our ordinary ways of making sense of action run dry.

In the remainder of this chapter, we trace two conceptions of 'will' as an active principle in deliberate self-movement within a classical understanding of rationality. The first, *sheer voluntarism*, understands the non-hydraulic activity of the will as what we might call 'sheer willing'—it breaks a deliberative impasse, but not a normative tie. The second, *normative voluntarism*, insists instead that the fundamental activity of the will is to undo normative underdetermination itself, rendering previously-underdetermined choice contexts rationally determinate. Each has a long history as well as a number of contemporary standard-bearers. Both are, I shall suggest, fundamentally flawed. But each fails in a beautiful, complex way—leading naturally to an expansion and radicalization of the voluntarist impulse.

3.2 Sheer voluntarism

Among the medieval scholastics, it was the Franciscans who most vocally endorsed hybrid voluntarism as a solution to the problem of free will and moral attributability within the classical rationalist framework. Peter John Olivi was perhaps the earliest, and while he was quite influential of the thought of later Franciscans like John Duns Scotus and William of

Ockham, he was rarely cited by name because his work was condemned shortly after his death in 1298.[86] Here is Olivi:

> When there is some number of equivalent things that are equally useful, nothing explains the will's adoption of one or the other of them except the freedom by which one is equally able to do this or that. Suppose there are two pieces of fruit or two people that are in every way and through all things similar and equivalent. Nevertheless, the will attaches itself to one of the two and leaves the other… [and] we manifestly feel that we do this from freedom of the will alone and not from some greater satisfaction in the one as opposed to the other.[87]

In this passage, Olivi appeals to the phenomenology of underdetermined choice-making to substantiate his claim that freedom in willing can be found in the 'liberty of indifference' with which underdetermined choices presents us. Elsewhere, however, he cites the importance of a capacity for willful choice in the face of normative underdetermination in his account of the value of friendship—in this case arguing from conceptual desiderata rather than phenomenological data.[88] And Scotus, more commonly cited as originating of the Franciscan notion of *indifferentia*, is insistent that choice in the face of rational indifference is essential to human freedom. His famous distinction between two fundamental inclinations in the will is meant to make morality possible by securing the

---

[86] Burr 1976. We know for a fact that Scotus read Olivi, since in one of his discussions of the Trinity, Scotus writes, "The third opinion is that of Peter John, which I will not report, due to the circumstances" (*Opera Omnia*, vol. I, d. 26, n. 42).

[87] Olivi, *Quaestiones in secundum librum Sententiarum*, vol. 2, q. 57.

[88] In a slogan: true friendship requires that a friend be chosen for her own sake, and not because there is more reason to choose her than another—which requires both rational underdetermination and the capacity of the will to move us in one direction rather than another without the perception of the chosen option as rationally favored above all others available. For an extended discussion of Olivi's views on friendship, see Kaye 2004. Related ideas on the contemporary scene include Badhwar 1987, Hursthouse 1991, Kraut 1986.

will's freedom,[89] making space for the human agents to choose freely between rationally-eligible options, without being spurred to the choice by any external or internal rational pressure:

> [If the will were] merely intellective appetite it would actually be inclined in the highest degree to the greatest intelligible good. But [since] the will is free, it can control itself in eliciting its act so that it does not follow its inclination, either with respect to the substance of the act, or with respect to the intensity, to which the potency is naturally inclined.[90]

Finally, we find that Ockham also cleaves to the liberty of indifference as the source and ground of the human freedom required for true moral responsibility:

> The will is freely able to will something and not to will it… To deny every agent this equal or contrary power is to destroy every praise and blame, every council and deliberation, every freedom of the will [*omnem libertatem voluntatis*]. Indeed, without it the will would not make a human being free any more than appetite does an ass…[91]

The form of willing the Franciscans seem to have in mind is what we might call *sheer willing*: volition carries the agent forward in conditions of rational indifference, when reason alone cannot. The exercise of will does nothing to alter or resolve the original normative

---

[89] See his *Opera omnia*, vol. 1, d. 39, qq. 5, 15; and vol. 3, d. 3, qq. 1, 22. See also comments on *indifferentia* in *Quaestiones Quoadlibetales*, q. 18, n. 9; qq. 16–17. (The idea was, of course, Anselm's before it was Scotus's: see his *De libertate arbitrii* [3]).

[90] Scotus, *Ordinatio,* vol. 2, d. 6, q. 2, n. 8 (W XII:353-4). While the will can direct itself towards either of a pair of contradictories, the intellect cannot, for it is determined to incline toward the good and not the bad. In Scotus's terminology, according to which natural powers incline in one direction while rational powers move freely, this (somewhat perversely, to modern ears) makes the will a rational power and the intellect a natural one. See the *Quaestiones subtilissimae super libros Metaphysicorum Aristotelis* d. 9, q. 15, nn. 4–7 (W VII:609–611); for further discussion see Williams 1995, esp. n. 6.

[91] Ockham, *Opera philosophica*, p. 321. For discussion, see McCord Adams 1987.

underdetermination; it breaks the deliberative impasse, not the normative bind itself. Importantly, these exercises of will are triggered only by conditions of underdetermination: when things are not indifferent to it, the will is inexorably inclined toward the best available option.[92] For this reason, the scope of volitional freedom is directly proportional to the prevalence of underdetermination in human experience, and the nature of this freedom is settled by the context in which it is deployed: it serves to propel action when intellection cannot do the job itself, because the available materials of choice are insufficient to build a rational bridge all the way to action.

Identifying 'liberty of indifference' as the *ur*-form of free willing is a tendency which characterizes the thought of a later group of Christian thinkers, the British moralists of the early modern period. Concerned more directly than some of the Schoolmen with articulating the complex of normative relations which constitute the moral domain, thinkers like Richard Cumberland, Joseph Butler, and Ralph Cudworth also made special space in their account of rational choice for a faculty of willing which, paired with the classical model, was limited to impasse-breaking in the face of normative underdetermination. Moral conduct is active, not passive; moral blame is properly directed not at a failure of theoretical reason (a misperception of the relative goodness and badness of available options) but at a distinctly practical failure; moral responsibility requires a form of self-determination which rules out a purely hydraulic theory of choice.[93]

---

[92] Actually, there is reason to think that Scotus thought that the will's freedom extended even farther than this: some of his comments (see esp. *Quaestiones Quoadlibetales*, q. 18, n. 9; q. 17) suggest that he believes that the will can incline toward even toward an option identified as 'lesser' than alternatives.

[93] Darwall 1995.

Cudworth is a good exemplar for this general tendency.[94] Like his contemporaries, he insisted on the necessity of a turn to the practical in our moral philosophy:

> [T]he Anticipations of Morality [do not] spring meerly from Intellectual Forms and notional Idea's of the Mind, or from certain Rules or Propositions, arbitrarily imprinted upon the Soul as upon a Book, but from some other more inward, and vital Principle, in intellectual Beings… whereby they have a natural Determination in them to do some Things, and to avoid others, which could not be, if they were meer naked and Passive things.[95]

This "vital Principle," distinct from the intellect and also from natural passions, is what makes adherence to moral principles possible. The normativity of moral law derives not from external imposition (by God or by some earthly sovereign),[96] nor from the intrinsic natures of things themselves, nor—as the passage above makes clear—from their status as modes of theoretical reason. These principles are and must be purely practical. Nothing arising from any of these other sources could have the right kind of authority, and so there must be such a thing in the mind as "determinate volitions or motives," as pure practical

---

[94] Cudworth, *Treatise on Eternal and Immutable Morality*, IV.v.4; p. 118. Cudworth's work is extensively quoted in Raphael 1991, vol. 1, pp. 105–119. References to the *Treatise* in this work include page numbers to that volume. References to his unpublished work refer to the source in Darwall 1995. See n. 20.

[95] I draw my analysis of Cudworth from the exquisitely-researched Darwall 1995. Darwall's analysis of Cudworth's work, especially his account of the will, draws on a collection of unpublished manuscripts which are not commonly available or consulted, which means that his work departs in significant ways from the treatment of Cudworth's thought elsewhere. For a more traditional but no less rich account of Cudworth's moral philosophy, situating it among that of the other Cambridge Platonists, see Hutton 2007. For more on the Platonic Renaissance in England, see Cassirer 1953, Patrides 1969.

[96] "Whereas it is not the meer Will and Pleasure of him that commandeth, that obligeth to do Positive things commanded, but the Intellectual Nature of him that is commanded" (Cudworth, *Treatise on Eternal and Immutable Morality*, I.ii.4; p. 109).

reason, which alone enables moral normativity and (concomitantly) genuine moral compliance.[97]

Yet Cudworth, unlike many contemporaries, worried about sheer voluntarism as a model for this crucial "active Principle" in the soul. While he insisted that truly free practical reasoning is necessary for the self-determination that makes morality possible, he worried about divorcing the faculty of will so fully from the intellect, such that its crucial operation was rationally unmediated and therefore, in his view, unintelligible. Will is not an independent principle of movement, which engages when the movement enabled by intellection of rationally determinate choices is ruled out; rather, it is a "reduplicate self-active principle" which unites the whole self in choice. It is as absurd to suppose that there is a faculty of will that wills as it is to think that, say, there is a faculty of walking that walks or a faculty of speaking that speaks; rather, it is the *whole person* who wills, and the *whole person* who is held morally responsible for her actions—including her judgments about what matters in life, about what she has reason to do and pursue.[98] Sheer voluntarism, even if restricted to those cases where reasons-based guidance has evaporated, is inherently unsuited to capture this idea:

> [I]f the blind will … remains indifferent to follow the last dictate of [the understanding] or not, and doth fortuitously determine itself either in compliance

---

[97] Gill 2004; *cf*. Darwall 1995 pp. 127–128.

[98] "[T]o attribute the act of intellection and perception to the faculty of understanding, and acts of volition to the faculty of will, or to say that it is the understanding that understandeth, and the will that willeth. This is all one as if one should say that the faculty of walking walketh, and the faculty of speaking speaketh" (Cudworth, *A Treatise of Freewill*, p. 24, as quoted in Darwall 1995, p. 132, n. 45).

> with the same or otherwise, then will liberty of will be mere irrationality, and madness itself acting or determining all human actions.[99]

The balance of reasons gives our choices meaning and content, structure and intelligibility. Sheer voluntarism is tantamount to the claim that there is a faculty of willing that wills—there is nothing else for willing to *be*, if we restrict its operation in this way. Cudworth worries that if we look to sheer will as a source of value in the human experience, as capturing those essential aspects of freedom, spontaneity, creativity, integrity, and moral attributability for which the hydraulic model leaves no space, we are bound for disappointment.

3.3 Problems for the view

As we've seen, the descriptions on offer for just what it is about the human experience that classical hydraulicism rules out are many and varied: the possibility of first warranting and then accepting divine grace; a capacity for moral responsibility; the pure autonomy of self-forging choice; the development of our individual character. Our interest here is not in whether any one specific proposal (say, Scotus's addition of the *affectio iustitiae* to the *affectio commodi* as a distinct mode of the will) is able to secure the good of human agency which it identifies as crucial (the possibility of adhering to a moral law in a manner unrelated to our natural inclination to pursue creaturely flourishing).[100] Rather, our goal is to specify the core of this cluster of notions—in §2.4, we called it 'the missing agent'—and to determine

---

[99] Cudworth, *A Treatise of Freewill*, p. 23, as quoted in Darwall 1995, pp. 132–133.

[100] For more on these two *affectiones* of the rational will, and their connection to Scotus's view of morality as a system essentially divorced from self-interest, see esp. Williams 1995. Williams' interpretation is somewhat unorthodox; for a contrast see Wolter 1997, pp. 31–42.

whether either of the modes of volition described in this chapter can expand the classical theory of choice to make room for a notion of rational agency which does justice to the basic impulse. We ask whether the sheer voluntarist project of situating the core of human agency in conditions of rational underdetermination is a coherent one, or whether it is, instead, a doomed effort, having made "liberty of will … mere irrationality, and madness itself" the determinant of all putatively meaningful human actions, as Cudworth feared.[101]

On a first pass, Cudworth's concern might seem overblown. There is a quite natural sense in which normative freedom directly enables attributable, owned, individual-driven action: as Thomas Hobbes tells us, "The Greatest Liberty of Subjects, dependeth on the Silence of the Law."[102] When normative constraints on action fall away—in this case, the rules of a political system—we have the freedom to act under our own steam. As Hobbes puts it, "there the Subject hath the liberty to do, or forbeare, according to his own discretion."[103]

Clearly, the normative silence Hobbes had in mind was the silence of actual laws. But the point generalizes to other kinds of normative freedom. Consider Bernard Williams's famous critique of utilitarianism:

> It is absurd to demand of [an agent], when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions… to neglect the extent to which his projects and his decisions have to be seen as the actions and decisions

---

[101] Cudworth, *A Treatise of Freewill*, p. 23, as quoted in Darwall 1995, pp. 132–133.

[102] Hobbes, *Leviathan*, xxi, ¶18.

[103] Hobbes, *Leviathan*, xxi, ¶18.

> which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity.[104]

The idea of integrity in action that Williams claims is ruled out by classical consequentialism is analogous to Hobbes's freedom to "do or forbeare." When external normative constraints—especially those of domains typically thought to silence, outweigh, or overrule considerations from other normative domains—fall silent, reasons of our own can bubble up and have their say. Action undertaken in such circumstances is owned by and attributable to the agent who performs it precisely because the reasons which tipped the normative balance and are reflected in her choice are drawn from her own personality, projects, or plans. They are not due to any external or universal perspective on her choice. An authorial relationship to our choices secures the qualities of individuality and attributability that are the core of the hybrid voluntarist project.

Yet there is an important disanalogy between these instances of what we might call *domain-constrained* underdetermination—as when the political laws or moral rules fail to apply to a particular case (**inapplicability**), equally favor two or more outcomes (**indifference**), and so on—and the sort of *thoroughgoing* rational underdetermination on which the sheer voluntarists pin their hopes. We saw above how the value of domain-constrained normative silence is linked to the determination of choice by 'personal' reasons. But in cases where, by definition, *all* of our reasons have run out, there is no further normative substrate from which considerations can arise to fill the void. Where normative silence falls, we may sometimes fill it with our own voices. But the form that this 'voice' takes in the cases above—identity-based reasons—is precisely out of reach in conditions of true

[104] Williams and Smart 1973, pp. 116–117.

rational underdetermination. How can choice reflect identity when our freedom begins only where our ways of making sense of action run out?

The Cudworth-style challenge to sheer voluntarism has a strong parallel to a critique of libertarian freedom of the will commonly traced to to David Hume, namely, that the opposite of causal determinism is not freedom but randomness.[105] The kind of freedom worth wanting—in our case, desired for the sake of individual attributability—can no more be found in the comparative unintelligibility of arbitrary choice than in the fully-rationally-determined (but fully-rationally-explicable) choices highlighted as the core of agency by the classicists. Rational arbitrariness cannot deliver a uniquely authorial relationship to our choices, any more than can hydraulicism about the process by which choice is made. We have still mislaid the agent.

3.4 Normative voluntarism

The sheer voluntarism of the Franciscans and British moralists contrasts sharply with a different hybrid view, one gestured at in the medieval context by Dominicans like St. Thomas Aquinas. Aquinas' treatment of the phenomenon of underdetermination also involves a classical orientation to rationally-determined choice, but cleaves more closely to

---

[105] See *A Treatise of Human Nature* §§2.3.1–2.3.2, *cf.* his *Enquiry Concerning Human Understanding* §8: "liberty, when opposed with necessity, not constraint, is the same thing with chance; which is universally allowed to have no existence." (§8, ln. 72). Just before this proclamation, he hypothesizes that the phenomenology of apparent **indifference** might be to blame for many misguided efforts at a libertarian theory of will: "The prevalence of the doctrine of liberty may be accounted for … [by] a false sensation or seeming indifference which we have, or may have, of liberty or indifference, in many of our actions … a want of determination, a certain looseness or indifference. […] We feel, that our actions are subject to our will, on most occasions; and imagine we feel, that the will itself is subject to nothing…" (§8, ln. 72, n. 1).

the hydraulic notion that choice requires *some* distinction in options even to proceed at all. Whereas the sheer voluntarists saw the will flinging the deliberator forward in the absence of rational guidance, the normative voluntarist sees it creating new normative material—placing a thumb on the perfectly-balanced or unreadable scale—thereby enabling the ordinary route from reflection on the balance of reasons to the performance of a fully-intelligible, rationally determinate choice.

Aquinas's considered view is complex, but themes of normative voluntarism emerge in a context as straightforward as an examination of the paradox of **buridan's ass** or **aristotle's dog**. He begins with the challenge that hydraulicism (which which his own intellectualism about the appetites of the soul apparently entails) seems to pose to human freedom:

> [I]f two things are absolutely equal, man is not moved to one more than to the other… Now, if that which is equally [choiceworthy] with something else cannot be chosen, much less can that be chosen which appears as less [choiceworthy]. Therefore if two or more things are available, of which one appears to be more [choiceworthy], it is impossible to choose any of the others. Therefore that which appears to hold the first place is chosen of necessity. But every act of choosing is in regard to something that seems in some way better. Therefore every choice is made necessarily.[106]

Aquinas' explanation of why this view of rational choice is not in natural opposition to human freedom relies on the fact that there is but one end which *must* be considered good: the "perfect good" of happiness, which we will necessarily. Any particular manifestation of this good can appear to us as choiceworthy or not, depending on context:

---

[106] *Summa Theologica,* IIa, q. 13, a. 6.

> If two things be proposed as equal under one aspect, nothing hinders us from considering in one of them some particular point of superiority, so that the will has a bent towards that one rather than towards the other.[107]

That which is considered *most* choiceworthy cannot but be pursued by the rational deliberator. But although the will is directed at the good in general, action itself can only be directed at particular goods, which will inevitably fall short—so we have nothing to fear from the apparent inescapability of willing what is best.[108] We are likely to encounter many situations in which our basically hydraulic volition will stall, requiring us to engage in a new process, considering entities under an aspect not originally considered, breaking our deliberative impasse by breaking the normative tie itself via the introduction of new material to the rational calculus. The passage suggests that there is a shift from the intellectual judgment of the weight of reason on either side, to a mode extending the original parameters of relevance: we invite new considerations onto the scale, and the

---

[107] *Summa Theologica,* IIa, q. 13, a. 6; just above, Aquinas explains that "in all particular goods, reason can consider an aspect of some good, and the lack of some good, which has the aspect of evil: and in this respect, it can apprehend any single one of such goods as to be chosen or to be avoided" (IIa, q. 13, a. 6).

[108] Put somewhat more Thomistically: although choice is matter of comparison, and comparisons are made by the intellect, the making of the choice itself is nevertheless an exercise of will, not reason. The will is intellectual appetite, not mere sense appetite, and as such is aimed at universals, not particulars. The plurality of particular goods means that the intellectual appetite will have a plurality of good objects available to it. *Summa Theologica,* IIa, q. 13, a. 1.

balance of reasons is tipped in one direction rather than another.[109] Paralysis is lifted and choice may proceed because new reasons have laid new track before us.

There is much else in Aquinas's work, especially his legal theory, which suggests a normative voluntarist bent.[110] But at any rate, the idea that the will's contribution to choice is to place a thumb on a barely-readable or perfectly-balanced rational scale, rather than simply to fling an agent forward by disregarding the scale entirely, is a view with some contemporary adherents whose accounts of the creative power of willing are more detailed. They seize the power of willing that Al-Ghazâlî granted to God alone and put it to human use, insisting that this power—not sheer but *creative* volition—is the seat of crucial facets of human agency, including moral responsibility, but also non-moral authorial responsibility for individual acts and for one's life as a whole.

---

[109] As Thomas Williams has observed, the Thomistic means of securing freedom of the will seems to be a *diachronic* liberty of indifference: while the intellect presents *x* to me now as good, it might present *y* to me as good at a different time. In each case I am impelled to will what is best, but since it's not always *x*, the will is free. The Franciscans, by contrast, insist on the necessity of *synchronic* liberty of indifference: at the moment of choice, the intellect presents both *x* and *y* to me as good in the same choice setting, and I may will either, as my volition tends in the moment of choice itself. If there is always *something* we must will, then the will is not free in the sense needed for morality Williams 2011; see also Gallagher 1991.

[110] Briefly: for Aquinas, to command is an act of reason, rather than an act of will, though it is preceded by an act of will. This is true for God as much as it is for humans, who are faced—unlike God—with the task of creaturely self-governance, individual as well as social. What is the relation of human law to the natural law, which was laid down by God? It could be either a a deduction from or a determination of the natural law, since natural law does not manifest the specificity required to be a basis of common action. This specification, then, sounds like the creation of new normative material out of a normatively underdetermined context. I am grateful to Mark Murphy for discussion of this point. For more on Aquinas's legal theory and its connection to practical rationality, see Murphy 2001; *cf.* Richardson 1997 on specification in an Aristotelian, rather than Thomistic, contest.

On the contemporary stage, Robert Nozick explores (but does not ultimately endorse) a position of this kind in his *Philosophical Explanations*; Ruth Chang definitively espouses it in a series of articles on what she calls 'hierarchical' voluntarism about normativity and the will.[111] The two views are interestingly distinct, but I will focus on their major area of overlap, exemplified in Nozick's account of non-arbitrary choice, and Chang's discussion of rational identity.

Appealing at once to the phenomenology of underdetermined choice and to the conceptual desiderata of freedom, Nozick speculates that, perhaps,

> [Practical] reasons do not come with previously given precisely specified weights; the decision process is not one of discovering such precise weights but of assigning them. The process not only weighs reasons, it (also) weights them. At least, so it sometimes feels. This process of weighting may focus narrowly, or involve considering or deciding what sort of person one wishes to be, what sort of life one wishes to lead.[112]

Choices of this kind contain an essential reference to the identity they partially create; Nozick speculates that the "self-subsuming" nature of such choices might render them non-arbitrary. Further, since they implicitly involve adopting a new policy about the relative weight of certain reasons for the sake of making a rationally-explicable choice, they set up a normative framework within which future choices can be made.[113]

---

[111] Some of her earliest work in this vein calls her view not 'hierarchical' but 'hybrid voluntarism,' the description to which this chapter owes its name. My work in this chapter has been very deeply influenced by conversations with Ruth about her work and related points, for which I am grateful.

[112] Nozick 1981, p. 294.

[113] Nozick 1981, p. 297.

> Another way in which bestowal of weight upon reasons can be nonarbitrary is that the self can synthesize itself around this bestowing: 'I value things in this way.' If in that reflexive self-reference, the I synthesizes itself (in part) around the act of bestowing weight on reasons, then it will not be arbitrary or random that *that* self bestowed those weights.[114]

Because this shift in policy takes place in the course of deliberation and is solidified by taking the action dictated by the newly-forged normative principle, these instances of normative creation are an important means of self-constitution. Our policies of treating certain kinds of favoring considerations as weightier than others in large part comprise our practical identities: I treat considerations of the form "*x* would be elegant" as weightier than those of the form "*x* would be humorous;" your ranking is the obverse; we are each rationally permitted to judge and act as we do; your cheeky sense of humor constitutes an important part of your individuality, and my admiration of grace and panache a central part of mine. When I make a choice that forges a new policy, the *self* that moves forward from that choice has been altered because the constellation of policies that comprise it has changed.

A similar connection between self-forging and underdetermined choice is central to Ruth Chang's 'hierarchical' hybrid voluntarism. On Chang's view, rationally-underdetermined choice-making has two stages. First, we assess the existing balance of reasons, and discover that some variety of underdetermination is afoot—**indifference**, **indeterminacy**, or the like. Then the will steps in: but on Chang's view, the proper role of the will is not to fling the deliberator towards an available alternative, but rather "to will a reason that supports

[114] Nozick 1981, p. 306.

an alternative."[115] This exercise of will "creates normativity by creating new reasons whose normativity derives from the very act of will," which is itself "part of a deliberative process of making oneself into the distinctive rational agent that one is."[116]

Only by creating normative material to bridge the gap between deliberation and action in underdetermined contexts can we choose in a way that is at once reasons-responsive *and* self-governed. The standard relationship between the option selected and the balance of reasons prior to choice is established by a creative act of willing; the choice is thereby intelligible and non-arbitrary even as it is genuinely free. The making and unmaking of such choices constitutes "the governing of our ideal rational selves," which is "arguably the central—and most exalted—exercise of rational agency."[117] No merely hydraulic picture of rational choice could ever capture this exercise of a crucial, volitional, and uniquely human capacity.[118]

---

[115] Chang 2009, p. 255.

[116] Chang 2009, p. 255: "…the proper role of the will in the second stage of deliberation is not to will an alternative but to will a reason that supports an alternative. This willing creates normativity by creating new reasons whose normativity derives from the very act of will. …creating reasons through an act of will is part of a deliberative process of making oneself into the distinctive rational agent that one is."

[117] Chang 2009, pp. 261–262; *cf*. p. 259: "Unlike the sheer willing of an alternative, willing a consideration to be a reason is part of the process of making oneself into a distinctive normative agent, that is, creating one's own 'rational identity.'"

[118] Chang 2009, pp. 254–257; *cf*. pp. 261–262: "Most philosophers concerned with explaining 'self-governance' … focus on how we make ourselves into creatures with the foibles and peculiar irrationalities or personality that we each actually have. But […] we are authors not only of our actual lives but also of our ideal rational lives—of the best we can be, rationally speaking."

3.5 Problems for the view

Normative voluntarism seems at once promising and obscure. Lifting the obscurity and making good on the promise comes, according to proponents, from the envisaged connection between identity and willed choice. Below, we proceed through a series of natural puzzles about the view as we try out alternative ways of understanding the basic proposal, noting what each can and cannot claim for itself. In the end, we shall find that the death knell for normative voluntarism rings in the same tenor as that for dogmatic rationalism. Adherence to the hydraulic theory of will—even partial adherence—rules out a tenable theory of underdetermined choice-making, here as before.

We begin with the basic challenge: how could an exercise of will alter the very balance of reasons to which choice-making is ordinarily merely responsive? What does it mean to place a thumb on the rational scale? It will help to differentiate the proposal from two close cognates: *a.* the notion that extended deliberation will sometimes bring to light considerations relevant to choice but not earlier in view (recall Aquinas), *b.* the idea that developing or shifting a preference during the course of deliberation can give us new desire-based reasons for choice, thereby altering the balance of reasons.

a. Information-gathering is an important part of, or precursor to, deliberating about what to do. We gather information about the options at hand—this flight leaves early in the morning, this one has a vegan meal option, this one is a red-eye—as well as, where relevant, information about our own preferences or evaluative judgments—red-eyes are intolerable, airplane meals irrelevant, early-morning departures pleasant. The course of figuring out what there is most reason to do involves gathering this data and then assessing

where, exactly, the aggregate directs us. Often, when a choice seems underdetermined at a first pass (say, because there is no straightforward way to compare the merits of alternatives), further reflection leads us to a judgment that one alternative is, after all, optimal. Although there is a sense in which the normative balance has, from the deliberator's perspective, moved from indeterminate to determinate, no exercise of will was required. The course of deliberation in such cases does not *create* the reasons which tip the balance, it simply *discovers* them through additional research and reflection.

b. Closer to the voluntarist proposal is the idea that we can, through the course of reflection, come to desire certain options—or certain aspects of options—differently than we did prior to initiating that course of thought. Suppose that all reasons, at bottom, derive their normativity (and dimensions, like valence and weight) from our desires.[119] Then a change in desire *just is* a change in the balance of reasons. If deliberation in an underdetermined context triggers some change in our subjective motivational set, which leads us to make a choice we see as supported by the reasons as they now strike us, then we have directly created new normative material (or directly altered that which was already on hand).

The problem here is to articulate the way in which this 'change of heart' in the course of deliberation—from a set of desires which underdetermine the rationally-appropriate outcome to one which singles out one option as best—is meaningfully distinct from the situation in *a.* Has the agent, through some combination of introspection and imaginative projection, discovered that her desires tend more in one direction than she'd earlier

[119] Schroeder 2009, Smith 1994.

thought? Surely this often happens; and surely it is often being forced to choose in this way that teaches us new things about what we truly desire, what we actually admire, what we simply couldn't stand. Yet these are cases in which reflection has merely *brought to light* latent desires, or forced a comparison between competing desires which seemed to be of roughly equal strength, and turned out on closer examination not to be.

If, instead, the agent has *brought herself* to desire some option, or some attribute that the option manifests, more than she did before, thereby resolving the deliberative impasse she faces, then she will have created rather than merely uncovered new normative material relevant to choice. The question, of course, is how. 'Desiring at will' is a notoriously difficult enterprise, as everyone from William Shakespeare[120] to Francis Hutcheson[121] assures us. Suppose you shift your preferences—or your subjective valuation of the options at hand—such that you now face a different constellation of reasons. Once more the question is: have you discovered that something is, in fact, more valuable to you than you'd previously realized? Or have you made something more important by committing yourself to it more fervently than before? If the former, then we are back in the situation in *a.*, where no new normative material was created. If the latter, then what does this commitment consist in besides an exercise of will?

And so we've arrived at a crisper restatement of the first challenge for normative voluntarism: to understand just what it is about an exercise of will in underdetermined

---

[120] See esp. Sonnets CLIII–CLIV, in which a cool well meant to quench the heat of desire instead turns warm without altering or diminishing the original feelings. The poems are probably based on a poem in a 5th century anthology of Greek mythology attributed to Marcianus Scholasticus.

[121] "…neither benevolence nor any other affection or desire can be directly raised by volition," in Raphael 1991, p. 274.

circumstances that could enable it to alter the balance of reasons. Separating the issue from that of merely *discovering* new reasons, and then running the question through the two major models for explaining practical normativity—desire-based and value-based—gave us two flavors of the very same problem.[122] Normative voluntarism seems to require that the will interfere directly with the balance of reasons. But there is no sense to this notion unless we understand the will as interfering directly with the materials that *constitute* the reasons themselves. Reasons are *abstracta*; as units that can be aggregated and weighed, they stand in for the considerations that actually favor the outcomes in any given choice context. Yet most ways of understanding rational practical deliberation see the materials that constitute reasons—that ground their normativity—as necessarily beyond interference. Whether given desires or given values, they are not up to us, because if they were, *everything* would be and there could be no rational constraint on action at all, ever. The trick for the normative voluntarist is to explain, first, how these materials could ever be subject to our direct meddling, and second, how the authority we have to so meddle is an authority we *only sometimes* have.

---

[122] There is a third way of understanding practical normativity, and that is as seeing the normative authority of all that is normative as residing in the will. Thoroughgoing voluntarism about practical normativity lives in the contemporary scene as constructivism about reasons, which we will revisit in the next section (**§4**). But the normative voluntarist is not a thoroughgoing constructivist—she is a hybrid constructivist, who sees only some practical normativity as arising from the will, as owed to something beyond the ordinary source of practical normativity, be it subjective desires or objective values or some combination thereof.

The claimed answer comes from the envisaged connection between the balance of reasons and our practical identity.[123] Consider one way we ordinarily alter the normative landscape we inhabit: taking on commitments or making plans. When I promise you to meet you for lunch, I thereby acquire a whole host of reasons: a reason to be at the cafe at the appointed time, a reason to call you if I can't make it, a reason to avoid making other commitments that would conflict with our lunch date, etc. Probably there was existing reason (desire-based *or* value-based) to meet you for lunch, which helps explain why I made the commitment I did, though in most cases I might have acted on *that* reason by making a slightly different plan. To coordinate our behavior, and to pin down whatever other plans I have for the week, we needed to settle on some time and some place. Now that we have, we have each freely and voluntarily given ourselves new reasons for action.[124]

Arguably, the same pattern applies to other small-scale coordination problems, temporal rather than interpersonal. I am biking home and need to decide whether to take the M Street bridge or the P Street bridge. It is rush hour and M Street is likely to be cramped and unpredictable; there is construction on P Street which will make the ride there no less arduous. Though without good reason to choose one route over the other, I still need to settle on *some* plan to get home, and so I opt for P Street, dodging cement-mixers all the way across town. I formed an intention to pursue an option which the balance of reasons,

---

[123] Or identities, plural. I shall speak throughout, as our representative normative voluntarists do, of each person as having a practical identity, but this is not meant to beg the question against theories of personal identity and integrity which see the self as a nested or overlapping set of divergent identities, each more or less internally-unified but with no dominant identity. Many thinkers, including no doubt some normative voluntarists, see practical unity and integrity as essential to agency; a few have begun to contest this assumption. See, e.g., Sebo (ms), for an exemplar and a useful list of references to fellow travelers.

[124] Watson 2004.

offering no guidance, did not favor. According to some, I thereby gave myself a reason to ride down P Street, or at least to do so unless some new information shows me that the balance of reasons actually tips in favor of M Street after all. I 'bootstrapped'[125] a reason for action out of thin air, because I had reason to make a plan rather than deliberate indefinitely, and so was rationally permitted (perhaps required) to 'plump' for an alternative not already favored by the balance of reasons as I saw them.[126]

In the first case, that of promising, we have a way to explain the apparent creation, through an exercise of will, of new reasons for action. I have some existing reason to do the thing I have promised you to do, and a corollary reason to translate that reason into a determinate plan so that we may each settle our other appointments for the week. What is the source of the normative authority of the reasons we have given ourselves by making this promise? It is arguably *not* in whatever existing reasons we had to see each other for lunch; rather, it comes from the relations of accountability we have established with one another in making the plan. If another commitment comes up, only you can release me from the obligation to meet you at the promised time; I cannot release myself simply because I have recognized that the constellation of given reasons now directs me elsewhere. Depending on how important the new commitment is, the balance of reasons may still favor abandoning our plan, with many apologies. But the new balance of reasons contains, inescapably, a reason to meet you for lunch *because I promised you I would* (because you now expect me to be there,

---

[125] As in: pulling oneself up by one's own bootstraps, performing a seemingly impossible task of fashioning the force or materials needed to move forward out of nothing more than what one already has on hand. For discussion, see Herman 2002.

[126] Bratman 2000, 2007. My thoughts in this section are much indebted to Luca Ferrero and Elijah Millgram, for separate conversations on this and a series of related topics.

because you in turn made time for me). The evaporation-conditions on that reason are at your discretion, not mine.

The same is not true in the second case: if I see that construction is far worse than expected on P Street, and the balance of reasons definitely points, contra my original judgment, to taking the M Street bridge, my original plan ceases to have any meaningful normative standing and I can quickly alter my plans. When I change plans in this way, do I still have the ghost of the reason I supposedly gave myself in choosing P Street, weighing against the considerations that favor M Street now that the situation has changed? Does the original plan carry weight on its own once the (illusion of) underdetermination has been washed away? Surely not—what then is the sense in which it was ever a *reason* at all? It is a live question regardless of whether we consider the normative bootstrapping that comes with intention to be the creation of a first-order reason to do the thing intended,[127] a second-order reason not to consider alternatives unless prompted to do so by more information,[128] or both.[129] The evaporation-conditions on each are at my discretion, and so the binding force of the original reason(s) seems suspect.[130]

This is all by way of saying that 'bootstrapping' a reason into existence looks like the sort of thing the normative voluntarist has in view. As Barbara Herman puts it:

---

[127] Smith (ms).

[128] For the origin of the notion of exclusionary reasons, see Raz 1975; for their application to the question of future-directed intentions (though without direct attribution) see Rovane 1998, p. 144*ff*.; Hinchman 2003.

[129] Concerns about bootstrapping have moved some to argue that we account for the normative authority of intentions by appeal to normative requirements rather than reasons at all. See Broome 2001, *cf*. Brunero 2010.

[130] Ferrero 2010.

> The attraction of bootstrapping is that you use a bit of what you already have to get some place you haven't been before, but need to go. As a strategy of argument, it is environmentally neutral. No new resources—no new entities or capacities—are called for… The bootstrap is made out of the same plain materials, only attached in a different place, so as to provide additional elements of argument.[131]

The normative voluntarist wants a concept of creativity and self-forging that is not available on the hydraulic model of choice alone, and looks to the setting of normative underdetermination to find it. But because the will must *do* something other than fling the deliberator forward in a rationally-arbitrary direction, she needs some account of how the normative landscape has been shifted by the will to enable rational choice. An exercise of the will interferes with the normative materials which ordinarily constitute the balance of reasons—or with the balance itself—but does not replace them with anything spooky or essentially new. The choice's comparative intelligibility is the same as in the classical model, namely, conformity with the balance of reasons. These reasons, though volitionally-altered, act as reasons like any other, and so can play their accustomed role in the explanation and justification of choice. What the voluntarist needs is a sleight-of-hand, which—Herman explains, in her inimitable fashion—is just what the move involves: "Its ontological and theoretical abstemiousness make bootstrapping one of the more elegant modes of philosophical prestidigitation."[132]

If successful, this model—a rational bootstrapping akin to that of promising or plan-making—has the right sort of shape to secure the bounded power of the will that the

---

[131] Herman 2002, p. 253.

[132] Herman 2002, p. 253.

normative voluntarist needs.[133] But it must be a notion of intra-personal commitment that *really does* bootstrap new reasons into existence, one that can hold them in place in the face of manifest underdetermination. Accounts of intention which bootstrap new reasons out of thin air typically explain their normative status by appeal to the practical requirement of unity in agency across time, or some other diachronic organizing principle.[134] These are surely needs which, even if they *can* do the bootstrapping required, apply primarily to future-directed planning, and not to any old underdetermined choice.

Yet there are major similarities between the kind of normative creation that the normative voluntarist has in mind and the normative bootstrapping of commitments and plans described above. In each case, normative creation is supposed to be made possible by the connection between reasons and practical identity. The latter ties the normative legitimacy of intention-based reasons to self-governing policies which structure our practical identity and thereby secure our agential unity across time. These are policies about what sorts of considerations to count as reason for action, in a thin and somewhat formal sense; yet as we noted above, much richer aspects of our personal identities have the same sort of footprint in the practical domain: they are policies about how considerations of different kinds (elegance, expediency, humor, color) feature in our judgments about what to do. Where reasons fall silent, our character is forged by the choices we make to bridge—to fill—that normative void.

---

[133] In fact, Chang argues for precisely this model in "Commitment, Reasons, and the Will," forthcoming in *Oxford Studies in Metaethics* (2013).

[134] Bratman 2000, 2001; *cf.* Ferrero 2010.

The question, of course, is how these choices come to alter our practical identity. For 'will-created' reasons to play a role in self-forging, the self that carries on from the moment of choice must be marked by the new policy about the weight of reasons it adopted for the sake of moving forward in rational space. On the normative voluntarist view, rather than arbitrarily 'picking' one option over another, without any distinguishing reason for it, we have *chosen* it—but we have chosen it for reasons (or strength of reasons) which did not exist at the time when we conducted our first survey of the normative landscape. What we are owed is an account of the normative authority of the identity-based reasons we have bootstrapped into existence to span the gap.

Nozick proposes that the answer lies in the 'reflexive' structure of choices involving will-forged reasons. In identity-forging choice, part of what is required to make the choice context determinate is a more concrete sense of oneself and one's "appropriate life": the choice is not just of an option, but of a way of life (self-conception) that makes sense of that option, where the self-conception involves the particular bestowal of weights that makes the choice context determinate. In this way, the suggestion goes, the choice "subsumes and thereby explains itself."[135]

There are then, on the hybrid view, two aspects to this self-subsumption. First, the adopted policy about the assignation of weights is self-approving; second, perhaps as a

---

[135] Nozick 1981, pp. 300–301: "Consider a self-subsuming decision that bestows weights to reasons on the basis of a then chosen conception of oneself and one's appropriate life, a conception that includes bestowing those weights and choosing that conception (where the weights also yield choosing that self-conception). Such a self-subsuming decision will not be a random brute fact; it will be explained as an instance of the very conception and weights chosen. [...] It will no more be a random brute fact than is the holding of a fundamental deep explanatory law that subsumes and thereby explains itself."

result, the self "synthesizes itself" around the then-chosen self-conception that yields the distribution of weight to reasons that offers definitive rational guidance.[136] Let's take them each in turn.

The first references a feature of normative policies often thought to be crucial to their stability, authority, or legitimacy: reflexive self-approval. Consider the 'self-effacing' nature of hedonistic egoism or simple utilitarianism,[137] the optimizing aspect of the choice to adopt optimizing as a rational strategy,[138] the satisficing form of the choice instead to satisfice,[139] the egoism of rational egoism, the moral self-approval of the moral perspective.[140, 141] Or consider the rational dilemma of **kierkegaard's editor**, who is presented with a choice between two fundamentally different approaches to life: the

---

[136] Nozick 1981, p. 306.

[137] Parfit 1986, p. 43.

[138] Byron 1998; *cf*. Slote and Pettit 1984.

[139] Schmidtz 2004, p. 37.

[140] That the normativity of morality is its capacity for reflexive, reflective self-justification, is a theme which both Christine Korsgaard and Annette Baier have claimed to find in the work of David Hume, building on his assertion that the moral sense "must certainly acquire new force, when reflecting on itself, it approves of those principles, from whence it is deriv'd, and finds nothing but what is great and good in its rise and origin" (*A Treatise of Human Nature,* III.3.iv); *cf*. Korsgaard 1996a: "According to this theory a faculty's verdicts are normative if the faculty meets the following test: *when the faculty takes itself and its own operations for its object, it gives a positive verdict*" (p. 62, see also pp. 60–64); compare Baier 1985, esp. pp. 157–173.

[141] Another familiar example from the history of philosophy might be Henry Sidgwick's "dualism of practical reason" in his *Methods of Ethics*. Yet as Brink 1988 notes, it's hard to know whether the dualism of practical reason is that between two moral theories (ethical egoism and utilitarianism), between two rational theories (rational egoism and utilitarianism), or between theories of rationality and morality (rational egoism and utilitarianism) as practical perspectives generally. *Cf.* Frankena 1974.

'aesthetic' and the 'ethical,' each of which (we suppose) provides an internally self-consistent set of normative standards and ideals to guide choice:

kierkegaard's editor You must choose between competing, self-ratifying normative systems.

One of the things that seems to make the editor's choice so terrifying—or so absurd—is the apparent inescapability of whatever framework is selected. In order to choose at all, one must have a system of assigning weights to *some* reasons, yet the choice of a fundamental orientation to value seems to place the editor outside the reach of any justification for choosing either framework. Yet once the choice is made, there would seem to be no escape from the perspective adopted: each perspective approves of itself (and of the choice to adopt it), and any time the question of whether to shift orientations arises, it is quickly silenced. (Perhaps it is impossible even to raise the question 'ought I become an aesthete?' when one has committed oneself to an ethical life.) There is a *stickiness* to the choice of a perspective like this one that is, perhaps, a feature of these perspectives' reflexivity. We have been robbed of the very language to ask and answer the question of escape.[142]

Nozick's reference to the reflexive aspect of choices which connect policies about what to treat as a reason in this case directly to chosen self-conceptions seems to appeal to much

[142] This is, of course, only one way of understanding the case of **kierkegaard's editor**. According to one compelling competing analysis, what the 'dilemma' really shows is that the ethical perspective, being the one to which rational judgment basically belongs, governs his choice, and is the *truly* inescapable perspective. The aesthetic perspective, on this reading, is not so much a competing set of values and policies about what is worth pursuing and celebrating in life, as it is a rejection of any such policy in the first place. I am grateful to Bill Blattner for discussion on this and related points.

the same structural feature. The self *cannot but* synthesize itself around that perspective, in making the very choice itself—'I value things in this way'—and having made a choice on this basis, the 'I' who emerges must truly value things in this way, even if the 'I' who arrived at the choice initially did not. If it were not true at the moment of choice that the new 'I' valued things in this new way, the balance of reasons would not have been shifted at all.[143] The bootstrap is made out of the same materials we had on hand at the moment of choice, but they've been re-arranged and re-attached in a new place.

Yes, but *how*? There are some serious problems with this account of the normative stickiness of adopted self-conceptions. Suppose that the explanation of **kierkegaard's editor**'s choice above is right, and the choice between competing normative perspectives will be genuinely irrevocable due to the adopted perspective's inescapability. First, the irrevocability of a choice does not make it self-justifying, any more than brute facts are self-explaining.[144] Put another way: that the proposition 'things with feature *x* are worth pursuing' manifests feature *x* does not make it any more worthy of pursuit than a contrasting proposition which also falls under the class of things justified by itself—or one which does not have this reflexive feature at all. Yet rational justification in the form of comparative choiceworthiness is what we seek for the choice made; this justification is supposed to be needed to make it non-arbitrary. The dilemma of **kierkegaard's editor** is

---

[143] Nozick 1981, p. 306.

[144] Anscombe 1958. Brute facts are typically described, not as self-justifying, but as beyond justification. Not all brute facts (or even all brute laws of nature) could be self-subsuming, even if some interesting subset is. I am grateful to Richard Fry for extremely helpful discussion on these points.

not resolved by the fact that the editor is likely to be (normatively) satisfied with his framework moving forward; it is amplified by this fact.

Second, the irrevocability of the editor's choice is more plausibly due to the fact that what is being selected is an entire normative system than that these systems are self-reinforcing rather than self-effacing. When we choose a wholesale system of value rather than a piece of one, as in the choice between egoism and altruism, or between meta-rational strategies, or in the choice between satisficing and optimizing, the choice itself seems to place us beyond the reach of any existing system, since our commitment to one over another is precisely what is at question. That is what makes the editor's dilemma one of **inapplicability** rather than one of the other varieties of underdetermination. But why suppose that ordinary underdetermined choice will take this form, even occasionally?

Consider again the type of underdetermined choices we are looking at: the normative voluntarist looks to choice contexts where *both* world-given reasons (including moral reasons and those of other important domains) *and* the agent's personal reasons underdetermine the normatively appropriate outcome. For the voluntarist, while the world-given reasons are not subject to adjustment through an exercise of the will, our personal reasons are, *because* they are part of a system of self-governing policies about what sorts of things are worth pursuing and why. We adjust our practical self-conception in order to adjust the balance of reasons (or, we adjust the balance of reasons and thereby adjust our functional self-conception—it comes to the same thing). Were the choice underdetermined from the perspective of the moral, legal, prudential, etc., domains, but *not* the domain of 'personal' reasons, it would not be rationally underdetermined. The relevant

underdetermination, then—the underdetermination which it is up to us to resolve—must be traceable to some aspect of our existing self-conception, to some immaturity, indeterminacy, ambivalence, ambiguity, or fragmentation of the practical identity which we use to navigate the practical domain.

None of these ways that our practical identity can underdetermine choice amounts to a failure of *application*. Perhaps a deep-seated ambivalence about the relative value of considerations of politeness and clarity in communication is to blame; perhaps we have never before had to manage such a direct trade-off between honesty and tact; perhaps we are for the first time forming a policy which situates a new and complex form of value within the existing nexus of practical policies. None of these is a choice between wholesale normative frameworks. In most cases an adjustment, specification, or shift of one's existing policies—policies which typically *explicitly* balance one sort of consideration against another—will suffice for the kind of normative creation the voluntarist seeks. Far from being robbed of the language to ask and answer questions about foregone self-conceptions, such choices speak a language which incorporates many of the individual values and desires that comprise an evolving, maturing self.

It is possible that a choice may rub against some deep fracture in one's existing self-conception, or reveal that one has not one but several competing self-conceptions, each of which speaks strongly in favor of a different available course of action, but none of which is the 'core' self. In such cases, perhaps, the deliberator occupies the functional perspective of **kierkegaard's editor**: fully identified with neither, yet unable to proceed without implicitly endorsing one, with the odd structural result that the adopted self-conception

will itself "be explained as an instance of the very conception and weights [of reasons] chosen."[145] In fact, this is one way of interpreting the root of the dilemma of **sartre's student** as well.

sartre's student You must decide whether to join the French Resistance or to stay home to take care of your ailing mother.

Yet surely such choices are rare: fragmentation of this form, where the two selves are sufficiently well-rendered as to constitute competing wholesale normative frameworks, and there is no core self from which judgments adjudicating between the two selves (and thereby between the two possible distributions of the weights of reasons) can be made. The activity of forging one's "ideal rational self," of building up and reinforcing one's practical identity, must take a different form.

So the "self-subsuming" or "reflexive" structure of rationally underdetermined choice does not make it non-arbitrary. Nozick attempted to secure rational justification by reference to a structural feature of the kind of choice being made. It was a paradigm instance of bootstrapping: using a bit of what's already on hand to get someplace new, without appeal to anything ontologically suspicious. The exercise of will itself was not what somehow secured the non-arbitrariness of the choice, through a bit of normative magic; if the nature of the choice itself was such that self-justification was inescapable, then an arbitrary selection *could not but* create new normative material. The will could move the agent forward in the face of underdetermination in just the way envisioned by the sheer voluntarists, and the choice context would do the rest.

---

[145] Nozick 1981, pp. 300–301.

Since it cannot, we are left again with the original mystery: what about the connection between identity and the will makes willed choice in the face of normative underdetermination normatively creative and therefore non-arbitrary?

Perhaps it is not the choice context itself—some normative slight-of-hand pulled by the möbius strip of reflexivity—but something that happens *after* choice which makes it non-arbitrary. If we are indelibly marked by the choice made, then it will not, as Nozick says, "be arbitrary that *that* self bestowed those weights [and then made that choice]."[146] Certainly some choices alter our life in radical ways, such that it becomes nearly impossible to imagine one's life having gone another way; consider **sartre's student** reflecting on his choice to join the Resistance; consider your own choice of a career or life partner. Certainly these choices make us subject to new sorts of reasons related to the path chosen: the student's choice to care for his ailing mother makes him host to a new set of obligations and reasons related to that path and not the one foregone. And certainly there are many psychological mechanisms that might—and probably often do—cast our choices in a rosier light than they deserve, viewed retrospectively.[147] The need to see ourselves as rational, our choices as justified, our life stories as sensible—even inevitable—rather than contingent is a deep and abiding one. And it furnishes a connection between one's (subjective, malleable, diachronic) self-conception and the perceived normative status of a choice (justified by the balance of reasons at the time it was made).

---

[146] Nozick 1981, p. 306.

[147] Including, but not limited to: choice-supportive bias, consistency bias, egocentric bias, endowment effect, fading affect bias, false memory, hindsight bias, illusory correlation, loss aversion, outcome bias, post-purchase rationalization, rosy retrospection, subjective validation. See §**4.3** for further discussion.

Yet this basically existentialist idea—that we might have normatively justified ways of making sense of our choices retrospectively—*cannot* be what the normative voluntarist has in mind. It seems to combine a radical subjectivism about normative facts (*they are what I say they are*) with some kind of backwards causation (*even if I now say they are different than what I said they were before*). Neither of these is licensed.

First, radical subjectivism about normative facts is definitively ruled out by the hybrid or hierarchical nature of the view. For the hybrid voluntarist, some reasons are world-given, some will-created; the world-given reasons govern the normative status of choice absolutely; the will-created reasons do too, but only after we've put them in place. Radical subjectivism applies only to the will-given reasons—only those are as we say they are—and we are only entitled to say that they are as they say they are when underdetermination triggers the exercise of this kind of normative authority.

Second, any notion of comparative intelligibility that rests on a misperception or reconstitution of past normative facts is also beyond reach. Such normatively-significant misrecognition would involve explicit appeal to a retrospective, diachronic form of justification, rather than the synchronic form to which the classical model is wedded. For the hybrid voluntarist, self-synthesis must be completed prior to the moment of choice, because the standard of intelligibility the voluntarist aims to secure is one of synchronous justification of the choice by the balance of reasons. According to the classical rationalists, all that one ever needs to say about this balance is resolved prior to deliberation. According to the normative voluntarists, the balance can be altered during the course of deliberation, through an exercise of will, but the balance of reasons must still be shifted before the

choice itself is made. If it is not, the choice will be arbitrary according to the only rational standard under serious consideration—strong comparativism—that is to say, arbitrary relative to the reasons one has at the moment of choice.

The agent's contribution to the meaning of her own choice was supposed to take the form of an altered self-conception. By effecting some shift in the constellation of self-governing policies about the assignation of weight to reasons—policies about what's worth pursuing and why—she would thereby effect a shift in the normative terrain she faced in choice, making an initially-underdetermined choice into a rationally-determined (and therefore intelligible) one. But what authority she had to make such a shift went unexplained. If reasons for action, including personal reasons, are typically treated as static from the perspective of the deliberating agent—as it seems they must be, for the hydraulic operation of reasons on the will to proceed as it does in ordinary cases—why treat them as malleable when they happen to underdetermine the choice at hand? Plausibly, we have this authority because in such cases the underdetermination of the choice reflects a kind of immaturity, ambivalence, or indeterminacy in our operative self-conception; just as in the case of future-directed planning, we have the authority to treat an arbitrary decision as fixed for the purpose of organizing our other activities. In a sea of arbitrary organizational choices, something must be held steady, or the agent herself runs the risk of impotence or even disintegration. This justifies our treating arbitrary decisions as closed rather than open to continuous reconsideration.

In this latter case, we do seem to have 'bootstrapped' some kind of new normative material into existence in the face of rational arbitrariness, and are justified in treating the decision

made as fixed. Yet that decision is always open to reconsideration when circumstances shift —when we have reason to think that the underdetermination of the original choice context has evaporated, or was illusory all along. The staying power of the arbitrary choice is thin; the exercise of will that settles the decision does not alter the balance of reasons irrevocably, if it alters them at all (rather than effecting some more subtle normative shift). It must be, if these decisions are to represent exercises of ordinary rational agency rather than forms of self-deception or self-manipulation.[148]

By contrast, the substantive shifts in self-conception that the normative voluntarist are after—the ones that reflect alterations to one's practical identity, richly conceived—cannot be seen as essentially evanescent in this way. These are not future-directed choices taken to coordinate and execute plans; they are identity-driven choices where some aspect of one's personal integrity or individuality is at stake. The self-evident arbitrariness of the future-directed choices—arbitrariness relative to the balance of reasons as it was prior to deliberation, that is—leaves them open to rational reconsideration right up until the moment of execution arrives. Shifts in self-conception are not diachronic, for the hybrid voluntarist; they must be completed prior to the execution of choice in order to do the normative work they have been hauled in to perform. This openness to revision, it would seem, must similarly perdure up until the moment of choice—and when that choice arrives, what reason do we have to treat the new bit of identity we arbitrarily adopted (having had no more reason to shift our self-conception in one direction rather than another, when many would have sufficed to place a thumb on the rational scale) as more truly ours than the one with which we arrived at the choice itself?

---

[148] Ferrero 2010.

The hybrid voluntarist needs normatively-creative choice to set up "a framework within which we make future decisions."[149] But why should this be so, if his account of how we evade the arbitrariness of such choices fails? Having chosen to bike down P Street this time gives me no more reason to select it next time than anything else. If I had a great time on the ride, *that* might give me (or be evidence of) a reason to adopt a policy of traveling on that route during rush hour; if it was dreadful, I might next time have more reason to try M Street. In either case, it's what I learn from the world after I make the choice that gives me a reason, not the exercise of will that hastened me down one street over the other. Perhaps a further policy of 'seeing more of the city' gives me a reason to alternate my route choices; perhaps a meta-policy of cultivated spontaneity gives me reason to randomize them. Here again, even though these personal policies make the rational status of future choices depend on the choice I made last time, the source of whatever reasons I have to choose routes in the future is not any automatically-justified exercise of will in adopting a policy, but something outside of the original choice. Why think that a choice of policies when it comes to something closer to the heart of my practical identity should be any more fixed? Indeed, why should it not be even *more* in need of external justification?

We see, now, why the attempt to mirror the intention-based bootstrapping of new reasons was misguided all along. The bootstrapping strategy depends on having identified some essential connection between the exigencies of practical deliberation—something like the diachronic division of deliberative labor[150]—and the need to treat arbitrary future-directed choices as settled rather than open to further deliberation. Because the normative

[149] Nozick 1981, p. 297.

[150] Ferrero 2010.

voluntarist is concerned with present-directed decisions about matters which impinge, directly or indirectly, heavily or lightly, on our personal self-conception, appeal to some agential need to make some choice, any choice, and *treat it* as rationally-justified will not suffice. The materials we have on hand at the moment of choice are not up to the task of genuine normative creation, and so the bootstrapping move fails.

3.6 Conclusion

Is there any avenue left to the hybrid voluntarist? The aim was to locate the agent amidst the sea of rational forces envisioned by the classical model—and the proposed solution was to find agency in the space where rational pressure gives out but choice proceeds anyway. The sort of agency worth discovering was thought to be one which expressed an agent's individuality, which made her choices and actions attributable to *her* rather than to forces—rational or otherwise—operating within her. To do so, the hybrid voluntarist aimed to find a way for such choices to be at once rationally underdetermined, leaving some work for the agent herself to do, yet still self-intelligible. The form of intelligibility available on the classical model is justification by the synchronous balance of reasons, and so this was what the hybrid voluntarist aimed to secure.

Sheer voluntarism, we quickly discovered, never really stood a chance of securing the intelligibility required for the attribution of choice to the 'agent,' richly conceived. Both varieties of voluntarism fared much better than even Leibniz's dogmatism at the task of explaining the phenomenology of underdetermined choice; more importantly, each variety represented an important stab at some crucial conceptual desiderata about the nature of

free and meaningful human choice. Normative voluntarism built directly on sheer voluntarism's failure to find a way to wring meaning from those choices which seem to take place where our ordinary ways of accounting for the meaning of choice run dry. Yet it seems that in the end all it managed to do was to deepen the mystery of volitional activity.

When we cannot get this justification from the peculiar structure of the choice, we have nowhere to locate it but the exercise of will itself. And so we have only deepened the mystery, for the hybrid voluntarist is left with something as thin as al-Ghazâlî's assertion that an exercise of divine will could make the rational difference between two or more identical options—such as a discrimination among instants in which to create the world, or between the options to color an atom white or black—but not in any way intelligible to creatures like us:

> The imponderable decisions of God cannot be weighed by the scales of [human] reason.[151]

Yet the scales of human reason are precisely what we must employ to explain underdetermined choice, if it is to be attributable to its author in the way sought. Placing the explanation of underdetermined choice beyond this reach, we have officially given up on the project of locating the nexus of freedom and rationality, creativity and integrity, spontaneity and intelligibility, which were supposed to constitute a central exercise of human agency.

---

151 From *The Incoherence of the Philosophers*, as quoted in Averroes 1954, Introduction, p. x. For his discussion of whether an atom shall receive whiteness or blackness according to the will of God, see Fakhry 1957, p. 136.

Reflections of this kind seem to press us ever further towards the more radical conception of underdetermined choice with which we flirted above. The classical rationalists' tight embrace of hydraulicism left no space for a plausible theory of rational agency; the hybrid voluntarists' partial embrace proved no less deadly, for it precluded any meaningful account of non-hydraulic choice because of its inability to articulate a notion of intelligibility that was not beholden to the normative hydraulicism of paradigmatically rationally-determined choice. From here, the only move would seem to be to let hydraulicism go altogether—to decouple our accounts of comparative intelligibility and rational determination, accepting that the project of being human consists in more than responsiveness to sums of rational forces, involves recognizing that, as Gabriel Marcel writes, "whatever its ultimate meaning, the universe into which we have been thrown cannot satisfy our reason—let us have the courage to admit it once and for all."[152]

[152] Marcel 1956.

4 Radical voluntarism

Pattern-seeking readers will have noticed that the named cases of rational underdetermination used to introduce the concept back in §1 could be divided not just conceptually, but also historically, into two groups. To the first belong the characters who've played a starring role so far:

buridan's ass	Standing equidistant between two identical bales of hay, you are hungry and want a bite to eat.
aristotle's dog	Standing equidistant between equal volumes of food and water, you are as hungry as you are thirsty, and darn thirsty at that.
leibniz's god	You have an infinite number of possible worlds you might choose to create, and an infinite number of instants in which to create them.

Aristotle, Buridan, Leibniz: representatives of the classical, medieval, and early modern periods, in which the classical model of choice reigned supreme and a presumption of top-to-bottom rational intelligibility in the natural order frayed only along the very edges of

philosophical discourse. The kind of underdetermination that most thinkers who cared to discuss the phenomenon had in mind was essentially mathematical. It was due to some oddity of the choice context (multiple perfectly equivalent options), not to any basic problem with the rational structure of the world or the power of human reason to assess it. **Indifference** and a very particular form of **indeterminacy** have been the major forms of underdetermination countenanced so far in discussion, even contemporary discussion which still hews, if not to the dogmatic implication that other forms of underdetermination are impossible, to those aspects of the classical understanding of rational choice which make cases like **buridan's ass** seem like the *ur*-form of normative silence to which all other varieties may be easily assimilated.

Now compare the remaining two varieties of underdetermination, which bear the names, not coincidentally, of two philosophers representative of a modern (or postmodern) turn:

kierkegaard's editor	You must choose between competing, self-ratifying normative systems.
sartre's student	You must decide whether to join the French Resistance or to stay home to take care of your ailing mother.

The underdetermination in these cases—instances of **inapplicability** and **incomparability**—might be described as indicative of a true breakdown of the standard model of rational comparison, rather than some basically nonthreatening oddity of a few exceptional choice contexts. It is not a formal aspect of the choice but some essential aspect of the options at hand, or their relationship to the choice-maker, which makes the choice rationally irresolvable. Some choices take place outside the bounds of comparative assessment;

sometimes, it seems, comparison is impossible; at others, the nature of the choice places us beyond the reach of rational assessment altogether, comparative or otherwise.

It is also no accident that a great deal rests on the choices of **sartre's student** and **kierkegaard's editor**, while the choices of the menagerie or divinity seem more suitably resolved by an arbitrary selection; nor is it an accident that the members of the classical and medieval crew are animal and divine, while those of existentialist class are all too human. The richness, complexity, and pathos of these recognizably human choices, where genuinely dissimilar options that will give shape to a whole life must be assessed, makes a stark contrast with the trivial paradoxes of the earlier set. The hybrid voluntarist turn to underdetermination was either empty or hopelessly mysterious, a fact which owes a great deal to the choice those thinkers made to situate our creative agency in contexts where reasons ran out in a basically mathematical way, leaving no real normative complexity for the will to engage. Viewed in this light, of course the effort was doomed to failure.

4.1 Radicalization

Why become a radical? This shift in focus, from the global to the specific, from the divine to the human, helps explain the radicalization in views of underdetermination that crops up historically in the early twentieth century, as well as the impetus many might feel while tracing the dialectic of rationality and agency through the lens of underdetermination to the present point. Staring into the abyss of underdetermination, those I'll call *radical voluntarists* laugh at the absurdity of the human condition—at self-conscious agents condemned to seek intelligibility and meaning in a world which too frequently has none to

offer—and see that the only way left to move is past the classical model into a new territory. In a world where existence precedes essence, and every choice is inescapably particular and individual, not just the materials of choice but the forms of comparative intelligibility we use to explain them must be human creations. Our agency resides not in occasionally fiddling with the given materials of normative justification, but with setting—always and everywhere—the terms of that justification.

The radical voluntarist does not celebrate underdetermination, but faces it grimly as the only possible locus of human agency—indeed, as a basic and pervasive element of the human condition—and in so doing give the will a bit more to work with than the hybrid voluntarists could muster. Whether their account of the meaning and importance of underdetermination brings to light a satisfactory account of rational agency is the guiding question of this chapter. Ultimately, I argue, we will find that the radical voluntarist model of underdetermined choice is the first to leave enough space for the long-sought 'agent' to make a genuine contribution to the meaning of her choice. It does so by throwing away appeals to the balance of reasons, traditionally understood, and introducing instead the only form of intelligibility and meaning available in an era of postmodern disintegration, one which faces up to the radical freedom of the human condition but gives up on any coherent account of rationality in agency as it does so.

The radical is, in one sense, a classicist at heart—all those canvassed so far form an arc along the line sketched by strong comparativism as the paradigm of intelligible agency. For the radical voluntarist, however, classical rationalism shows up more in the guise of an *inescapable perspective* on choice than as a literal model of human choice. This is because, for

the radical voluntarist, it is underdetermined choice that is in fact the paradigm instance of human agency, the site of free and meaningful choice. Though all of our choices are, in a certain sense or on a certain level, rationally underdetermined, we are at the same time saddled with a deep need—even an inescapable feature of the first-person deliberative perspective—to see our choices *as* justified in the classical sense, and therefore intelligible. We cannot but yearn to see our choices as supported by the balance of reasons; if we are paying attention to the world around us, we cannot but notice that the our choices never measures up to this demand, and never could. This is our very human plight.

In this chapter, we examine two ways to be a radical voluntarist about human agency and rational choice. The first, a position in logical space I call *heroic existentialism*, is drawn from themes we find in certain existentialist authors of the nineteenth and twentieth centuries. This view, like that of the dogmatic rationalists, is infused with a particular brand of rational metaphysics—indeed, one which is in the relevant sense the exact opposite, for it holds that the world as it is *only ever* presents us with underdetermined choice contexts. The second way to be a radical has rather more adherents on the contemporary scene: *existentialist constructivists* about practical reasons also formulate a theory of rational agency on the basis of a claim that the world as we find it is bare of the normative materials (reasons) we ineluctably seek in reflective choice-making. Our agency basically consists in the exercises of will that create the reasons that justify our choices.

In the end, I will suggest that radicalization is an understandable temptation, but ultimately no more promising an account of rational agency than those on offer in the last two chapters. The dogmatic rationalists tried to derive agency from rationality, the hybrid

voluntarists to paste the will onto a theory of choice which had no space for it. The radical voluntarist completes the slide to the other end of the spectrum, and attempts (in so many words) to derive a theory of rationality from agency. And just like the dogmatic classical proposal, its extremity renders it, ultimately, untenable.

4.2 Heroic existentialism

Existentialism is somewhat difficult to delineate, historically. This is so largely because, of the many often grouped under this heading, only two actively embraced the label.[153] For our part, we leave labels to the historians, and simply examine an account of rationally underdetermined choice that is suggested by some of the work of a few of those given the title of existentialist: Søren Kierkegaard, Jean-Paul Sartre, and Friedrich Nietzsche. It is a view of human agency that takes off from the perceived failure of the classical model—not just its hydraulicism but its very account of what makes choice intelligible—to grapple with the modern or postmodern condition, in which foundational underdetermination is a pervasive feature of human choice. The standards of intelligibility in choice, if any there be, must depart radically from the classical standards.

We are interested, then, in existentialism with a distinctly heroic flavor—not nihilist, not quietist, but one characterized by what Sartre calls a kind of "stern optimism" about the possibility for meaning and intelligibility in human choice (and therefore human life) in the unbearable absence of classical rational standards. The view of willing—of choice

[153] Those two were Simone de Beauvoir and Jean-Paul Sartre; though Kierkegaard sometimes uses the term 'existential' to describe his philosophy, he does not do so with any real systematicity (Cooper 1999).

generally—possible in this context is predicated on two inescapable facts. First, that we humans are bound to seek meaning and intelligibility, in ourselves and in the world. Second, that the world as we find it cannot supply this meaning. Condemned to live in a world empty of normative guidance yet equally condemned to crave and project it, being human is by turns an absurd and a glorious problem, solved (insofar as it ever can be) by each individual, not through reflection, but through a series of self-launching choices. The intelligibility of the launch resides not in the moments prior to choice but in the cast we place on them, retrospectively, as one in a series of basically meaningless movements whose only meaning is the one we construct for them.

The rationale for this view is basically metaphysical. Whereas the dogmatist saw human choice as one among many species of movement in a rationally-organized system, the existentialist sees choice taking place against a backdrop stripped of the materials of rational organization—and therefore of classical intelligibility and meaning. Consider this famous Nietzschean proclamation:

> God is dead. God remains dead. And we have killed him. How shall we comfort ourselves, the murderers of all murderers? [...] What water is there for us to clean ourselves? [...] Is not the greatness of this deed too great for us? *Must we ourselves not become gods simply to appear worthy of it?*[154]

Or this Sartrean *cri de couer*, explaining his dictum that existence precedes essence:

> What do we mean by saying that existence precedes essence? We mean that man first of all exists, encounters himself, surges up in the world—and defines himself afterwards. [...] There is no human nature, because there is no God to have a

[154] Nietzsche, *The Gay Science,* §125, emphasis mine; *cf.* §§108, 343; see also *Thus Spake Zarathustra*, prologue.

> conception of it. Man simply is. …he is what he wills, and as he conceives himself after already existing—as he wills to be after that leap towards existence. Man is nothing else but that which he makes of himself. That is the first principle of existentialism.[155]

We find ourselves in a world bereft of any of the tools—including comparativism about rational choice—we naturally want to seize to make sense of it. We may despair, or we may forge a new set of tools entirely.

Heroic existentialism can be recognized by its bravado, its self-set task of building meaning in a desolate place, and by the source of meaning it singles out: not human subjectivity, but human *will*.[156] It is our choices, fundamentally, that provide the material and the structure for a new kind of map for making our way in the world. But this map is not something consulted prior to choice, as the classical model would have it. It is something we construct after the fact, for the purpose of making sense of our lives rather than molding them to match an existing rational balance. The map is constrained by the choices we have actually made; while our freedom to choose is radical, our subjectivity is not, and our interpretive powers are bounded by the shape of the life we have lived.[157]

---

[155] Sartre 1946.

[156] I realized soon after adopting it that I actually owe the term 'heroic existentialist' to my old teacher Christine Korsgaard, though I use the term somewhat differently than she does. See Korsgaard 1996b, pp. 63–65. We may both have been inspired by this passage from Sartre: "Whereas the existentialist says that the coward makes himself cowardly, the hero makes himself heroic; and that there is always a possibility for the coward to give up cowardice and for the hero to stop being a hero. What counts is the total commitment, and it is not by a particular case or particular action that you are committed altogether" (Sartre 1946).

[157] Kierkegaard, *Concluding Unscientific Postscript*, esp. II.i–ii.

Indeed, according to Sartre, subjectivity itself imposes limits on the sorts of meaning we can impute to our choices:

> …man is, before all else, something which propels itself towards a future and is aware that it is doing so. […] Before that projection… nothing exists; not even in the heaven of intelligence: man will only attain existence when he is what he purposes to be. *Not, however, what he may wish to be.*[158]

Though we cannot transcend our own subjectivity—there is no sense to an objective balance of reasons—our interpretive powers are not limitless; we cannot give meaning to our lives in a manner wholly untethered from the reality of our choices themselves. The meaning we create comes from action and not from intellection.[159]

We can trace the movement from sheer to radical voluntarism in the existentialists. Our choices fling us forward blindly into a future we have not truly chosen, and we are bound to be dissatisfied with this fact. The existentialist need not eschew standard forms of rational justification out of a desire to be *outré*; indeed, those I'm calling heroic existentialists were insistent that our ordinary ways of understanding intelligible choice-making are classical in form. Our most fundamental choices—those of basic values, of frameworks for assessing choices or casting meaning on propositions—are criterion-constituting rather than -guided; the arbitrariness of these foundational "leaps of faith" taints all subsequent choices.[160] As Sartre notes, for any given choice, whatever existing

---

[158] Sartre 1946, emphasis mine.

[159] "Subjectivism means, on the one hand, the freedom of the individual subject and, on the other, that man cannot pass beyond human subjectivity. It is the latter which is the deeper meaning of existentialism," from Sartre 1946.

[160] Kierkegaard, *Fear and Trembling*, esp. III.iii and epilogue.

commitments or values shape the choice context will themselves have been chosen in a radically underdetermined, and therefore fundamentally arbitrary, way.[161] Everywhere we look, we seek order and meaning—yet find only the abyss of underdetermination.[162]

So what do we do with this tension between what the human intellect craves and what the world can offer? We will—and we fashion ourselves into something intelligible only after the fact, because we cannot help ourselves. As Nietzsche writes in his notes, "That my life has no aim is evident even from the accidental nature of its origin; that I can posit an aim for myself is another matter."[163] Sartre agrees:

> When we say that man chooses himself, we do mean that every one of us must choose himself… To choose between this or that is at the same time to affirm the value of that which is chosen; *for we are unable ever to choose the worse.* What we choose is always the better…[164]

Sartre goes further than this, even—affirming the inescapability of the classical perspective on what constitutes justified choice,[165] even as we must make manifestly unjustified choices. He insists that this perspective has a universal aspect:

---

[161] "…what is usually called my *will* is probably a manifestation of a prior and more spontaneous decision" (Sartre 1946).

[162] "Courage also slays dizziness at the edge of abysses: and where does man not stand at the edge of abysses? Is not seeing always—seeing abysses?" Nietzsche, *Thus Spake Zarathustra,* "On the Vision and the Riddle," III, 2.

[163] Nietzsche 1954, p. 40.

[164] Sartre 1946, emphasis mine.

[165] Compare Nietzsche: "the axioms of logic … are … a means for us to *create* reality," yet at the same time "rational thought is interpretation according to a scheme that we cannot throw off" (*Will to Power,* §§516, 522).

> …in choosing for himself [man] chooses for all men. For… of all the actions a man may take in order to create himself as he wills to be, there is not one which is not creative, at the same time, of an image of man such as he believes he ought to be.[166]

Our lives are ours to live and then to make sense of; while the value of your life is "nothing else but the sense that you choose,"[167] at least one cast we give our choices must be that of an implicit assessment—known to be unjustified!—that the option we chose was best.

What does this all mean for a theory of rational agency? Have we given up on rationality altogether, in favor of a theory of radically free agency? Many of those typically labeled existentialists may well hew to this position—Marcel, for example—but those I've singled out as exemplars for heroic existentialism are, I think, more interestingly read as offering an *alternative* theory of rational agency. We have already examined Kierkegaard's concern with criterion-constituting choices (like that of his eponymous **editor**), and Sartre's insistence on our power to create meaning in a normative vacuum through will and self-consciousness. For both, the classical model of choice and the comparativist standard of rationality are at once inescapable and impossible. The classical model fails because we find ourselves in a world where our most fundamental choices are doomed to underdetermination, through **inapplicability** or **incomparability** or both, and so that way of making our choices intelligible—the most natural way—is unavailable. But thinkers like Kierkegaard, Sartre, and Nietzsche can be read, at least in places, as suggesting an alternate

---

[166] Sartre 1946. Compare, later: "Man makes himself; he is not found ready-made; he makes himself by the choice of his morality, and he cannot but choose a morality, such is the pressure of circumstances upon him. We define man only in relation to his commitments…"

[167] Sartre 1946.

mode of understanding for our choices. Rather than giving up on the possibility of life's meaning to dwell in the empty and absurd, these proposals involve appeal to a diachronic criterion for rational intelligibility.

What is this diachronic criterion, exactly? The classical model insists that the rational status of choices—and so, derivatively, their intelligibility or meaning—consists in their relationship to a contemporaneous normative fact, *viz.*, what the balance of reasons favored at the moment of choice. The existentialist account of the intelligibility or meaning of particular choices—and, derivatively, their rational status—must reside instead in those choices' relationship to non-contemporaneous normative facts. One set of facts that might have the right sort of shape for the job are *narrative* or *trajectory-dependent facts*, propositions that hold of a particular choice but whose truth-makers live outside of the moment of choice itself.[168] If there is some sense to be made of a particular choice, not in virtue of the relationship it had to the balance of available reasons, but in virtue of the place it holds in an overall pattern of choice across time, the choice might have a kind of diachronic intelligibility notwithstanding its synchronic rational arbitrariness.

The appeal to narrativity as a normative criterion is suggested by Nietzsche's much-contested doctrine of 'eternal recurrence.'[169] The core idea—that we are doomed to revisit

---

[168] Jones 2008; Lloyd 2008.

[169] "What if some day or night a demon were to steal into your loneliest loneliness and say to you: 'This life as you now live it and have lived it you will have to live once again and innumerable times again; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unspeakably small or great in your life must return to you, all in the same succession and sequence… The eternal hourglass of existence is turned over again and again, and you with it speck of dust!'" (*The Gay Science*, §341). Compare *Thus Spake Zarathustra*, "The Convalescent," III, 13; *Ecce Homo*, III, 3; *The Will to Power*, §§1057, 1059–1066.

our lives over and over again—is woven through a few of his works, most notably *Thus Spake Zarathustra*.[170] However, it is never quite clear whether the idea is meant as actual metaphysical fact (whether religious satire or in earnest) or instead to serve as a kind of psychological test: could we feel satisfied with the prospect of endlessly reliving our own lives?[171] Most of us, Nietzsche suggests, could not:

> Would you not throw yourself down and gnash your teeth and curse [at the prospect of eternal recurrence]? Or have you once experienced a tremendous moment when you would have answered [the one who offers you this future]: 'You are a god, and never have I heard anything more divine.'[172]

The 'psychological satisfaction' interpretation of eternal recurrence looks to our response to the prospect of reliving "every pain and every joy and every thought and sigh and everything unspeakably small or great in your life" for an assessment of the greatness or smallness of our lives. [173] If we view recurrence not with horror but with joy, we have met the Nietzschean criterion for "high-spirited, alive, and world-affirming" greatness:

> My formula for human greatness is *amor fati*: that one wants to have nothing different, not forward, not backward, not in all eternity. Not merely to bear the

---

[170] Nietzsche, *Thus Spake Zarathustra*, "On the Vision and the Riddle," III, 2.

[171] Defenders of the cosmological interpretation include Danto 1965, ch. 7; Kaufmann 1974, ch. 11; Strong 1975 (see esp. p. 261). Defenders of the psychological interpretation include Nehamas 1985, ch. 5; Paphitis 2009.

[172] Nietzsche, *The Gay Science*, §341.

[173] Nehamas 1985, p. 127.

> necessary, still less to conceal it—all idealism is mendaciousness in the face of what is necessary—but to *love* it.[174]

The language here is that of affect—we *love* our lives as we have lived them, in all their wayward complexity. Elsewhere, however, Nietzsche knits the normative notions of *meaning* and *value* to a retrospective view of our choices:

> To become what one is, one must not have the faintest notion what one is. From this point of view *even the blunders of life have their own meaning and value*—the occasional side roads and wrong roads, the delays, 'modesties,' seriousness wasted on tasks that are remote from *the* task.[175]

It is this 'normative' reading of eternal recurrence that I propose as a component of the heroic existentialist's account of normative intelligibility. It resonates nicely with Sartre's claims that our lives, like paintings, cannot be judged until they are complete:

> No one can tell what the painting of tomorrow will be like; one cannot judge a painting until it is done. [...] We are in the same creative situation. ...when we are discussing a canvas by Picasso, we understand very well that *the composition became what it is at the time when he was painting it*, and that his works are part and parcel of his entire life.[176]

The conception of our lives as works of art—or as literature—echoes throughout the existentialist canon. The act of living a human life is one of creation and invention rather

---

[174] Nietzsche, *Ecce Homo,* II, 10; compare his praise in *Beyond Good and Evil* for "the ideal of the most high-spirited, alive, and world-affirming human being who has not only come to terms and learned to get along with whatever was and is, but who wants to have *what was and is* repeated into all eternity, shouting insatiably *da capo* [a musical term meaning 'again from the beginning'] not only to himself but to the entire play and spectacle..." (§56).

[175] Nietzsche, *Ecce Homo,* II, 9.

[176] Sartre 1946, emphasis mine.

than conformity and regulation.[177] There are no rational structures preceding our entrance onto the scene; but a life without this structure would be absurd. We must create this meaning for ourselves by fashioning our own rational guidance. Yet although "all our actions are altogether incomparably personal, unique, and infinitely individual,"[178] heroic existentialism is not a nihilistic or even relativistic position. We create meaning for ourselves within parameters settled by our subjectivity, out of the materials of past and future choices, not introspection alone.[179] This meaning comes from the pattern that our life's choices make as a whole, which we affirm or deny in its entirety.[180] This affirmation, not responsive *to* but constitutive *of* normative fact, is the standard for intelligibility in the abyss.

4.3 Problems for the view

The previous chapter examined the hybrid voluntarist effort to temper the threat to agency posed by underdetermination, an attempt that ultimately celebrated the freedom for self-

---

[177] See, e.g., Sartre's remark: "There is this in common between art and morality, that in both we have to do with creation and invention. We cannot decide *a priori* what it is that should be done" (Sartre 1946); *cf.* Alexander Nehamas' famous examination of Nietzsche's work through a similar lens in *Nietzsche: Life as Literature*, see especially ch. 5 (Nehamas 1985).

[178] Nietzsche, *The Gay Science*, §354.

[179] Nietzsche, *The Gay Science*, §17: "There is no trick which enables us to turn a poor virtue into a rich and overflowing one; but we can reinterpret its poverty into a necessity so that it no longer offends us when we see it and we no longer sulk at fate on its account"; *cf.* Sartre's insistence on the material of past choices and the pressure of intersubjectivity as constraints on our individual subjectivity (Sartre 1946).

[180] "Have you ever said Yes to a single joy? O my friends, then you have said Yes to all woe" (*Thus Spake Zarathustra*, "The Drunken Song," IV, 19). Compare, too: "A new will I teach to human beings: to *will* this way which humanity has walked blindly and to affirm it" (*Thus Spake Zarathustra*, "On the Afterworldly," I, 3).

directed volition it seemed to present. As we found, the only notion of volition made possible on this hybrid view—part hydraulic, part voluntaristic—was not so much *self*-directed as entirely *un*directed. Without any tenable connection to the sense of self with which an agent arrives at the moment of choice, such exercises of will were empty.

The radical voluntarist celebration of underdetermination is in many ways a grimmer affair, though the freedom it affirms is more radical. This is a celebration of underdetermination with eyes wide open; yet it does involve the same fear of arbitrariness and unintelligibility that we surmise underwrites the classical rationalist project. As Sartre notes, the existentialists are properly reproached "not after all, [for] our pessimism, but the sternness of our optimism" about the human capacity to seek and build meaning.[181]

This sternly optimistic project, though, flirts with paradox on all sides. Moving beyond the classical model of intelligibility in choice—leaving a classical conception of *reason* behind—is perhaps doomed to this fate. Heroic existentialism requires a diachronic theory of normative intelligibility, squaring the temporality of self-conscious action with that of personal identity to form a coherent whole. As the authors highlighted so far emphasize, there is a kind of circularity essential to any such account: the self is a process, not a state; we do not arrive at choice with a determinate self but must fashion one through choice itself; our identities are created and affirmed through deliberate exercises of will which take place at a particular point in time, yet our understanding of the deeper meaning and authenticity of those choices depends on a kind of retrospective affirmation of constellations of choices rather than individual points of light. When we leave classical

[181] Sartre 1946.

reason behind, perhaps some such dizziness is inevitable. But as we seek a theory of normative intelligibility, we must find something more than temporal paradox as we mine this vein of thought—something more than an insistence on the metaphysical necessity of a basically circular account of identity, action, and will.

Let's begin by being a bit clearer about what we mean by 'a new criterion of intelligibility.' We are after some notion of *normative* intelligibility, one which still sees the basic favoring relationship between considerations relevant to choice and the choice made as an essential constituent of the interpretability and meaning of the choice. The account should be robustly comparative, for on the existentialist model, the criterion of choice is still 'bestness' rather than some notion of sufficiency. So we seek an account that explains why the option selected was 'the thing to do'; but our account of this univocal favoring of one option above others must differ from the classical 'favored by the balance of reasons at the moment of choice.' We replace this conception with a new one that still might lay some claim to the label 'rational.' What we abandon is the *synchronicity* of the classical account of rational intelligibility, not its other core commitments, including its sense of what a 'reason' for choice is in the first place. The question is whether any of the views so far canvassed can make good on their claim to do so—and if so, what this validation means for an existentialist conception of agency.

To restate the problem about circularity: there are basically two aspects to the stern optimism of the heroic existentialist. The first is an insistence that we create meaning for our choices at the time that we make them, in the way suggested by Sartre's painting analogy, Nietzsche's conviction that we are always (or ought always strive towards)

becoming who we are, or Kierkegaard's notion of a leap of faith that carries us past uncertainty to firmer ground. This aspect suggests that the moment of choice secures its own intelligibility, through the inescapability of a then-adopted normative perspective we cannot but impute to the self who made the choice. This is, notice, basically the entirety of the normative voluntarist view, though it trades inescapability for reflexivity and moves the instant of creativity to the moment of choice rather than to any deliberation that precedes it.

The second aspect of the proposal is the one emphasized in §4.2, in which *retrospection* plays an essential role in constituting the meaning of the choice. The importance of the conceptual relation between any particular choice and the pattern of choices which precede and follow it in time, of the necessity of affirming these individual choices by way of affirming the entire pattern, are all themes picked out as supporting a retrospective or diachronic criterion for the normative intelligibility of choice. Our choices make sense only the rearview mirror, not prior to the sheer act of will which propels us into an uncertain future.

Now, as we saw, our selected historical existentialists seemed to want to affirm both these understandings of agency at once: choice is directed at synchronic as well as retrospective self-understanding. Yet these two would seem to be mutually exclusive. If the heroic propulsion of sheer willing could secure its own intelligibility or meaning, there would be no need for retrospection; if retrospection were the criterion, no meaning would be possible at the moment of choice.

So intelligibility cannot be an all-or-nothing contemporaneous feature of choice. One reason for seeing this brave new intelligibility as a shifting and possibly gradable property is the appeal to narrative criteria implied by the heroic analogy of artistry to action. Narrativity, as a property of sequences, involves, at minimum, a *beginning–middle–end* structure.[182] This structure must either be unavailable until our life itself has ended, or else take for its terminus the moving now of current experience.[183] On this view, it is not just we ourselves, but also the normative status of our individual choices, which is in a constant state of *becoming* rather than *being*.

Let us take **sartre's student** as an example, just so we have something concrete on which to hang these speculations:

sartre's student	You must decide whether to join the French Resistance or to stay home to take care of your ailing mother.

Sartre's assessment of his student's dilemma is as a case in which the student's values are "uncertain" and "too abstract to determine the particular, concrete case under consideration."[184]

---

[182] Though not everyone would agree: see esp. Velleman 2006 for a contrasting view and a helpful survey of alternative positions.

[183] We might also see our lives as a collection of mini-narratives, whose endpoints can exist in a relatively stable state, though determining whether these together form a unified narrative whole might throw their demarcation into further question.

[184] Sartre 1946.

Let us assume, for the sake of discussion, that either of these courses of action is morally permissible.[185] It is the student's *personal commitment* to the equally morally choiceworthy—or incomparably morally choiceworthy—values instantiated in the options at hand which must settle the question of what to do. Yet he looks within himself and finds no answer. Perhaps his life has been structured by a core commitment to the value of *loyalty*, but this case, in which loyalty to a loved one and loyalty to the cause of humanity itself are at issue, his existing commitment underspecifies the appropriate outcome. His policy must be sharpened or specified as favoring one *sort* of loyalty over another; yet since his existing policies do not, there is nothing he brings to the moment of choice which can answer the question of in which direction this specification should proceed. Perhaps he has weighed these forms of loyalty against one another in the past—or perhaps he conceptualizes the options, correctly, as manifesting two rather different moral values (say, love and honor, or fidelity and justice)—but his existing self-conception accords them equal weight in deliberation. Or perhaps there is some other immaturity, ambivalence, or indeterminacy in his self conception which renders his existing policies incapable of specifying one option as 'the thing to do,' as favored by the balance of identity-based reasons as they exist at the moment of choice itself.

At any rate, the sense of self the student brings to his dilemma cannot furnish the answer which the choice context demands it must. Sartre suggests that his student follow his

---

[185] We must, since otherwise this will not be a case of true underdetermination. Of course, for Sartre, morality itself is something whose parameters we define in the course of choosing, but let us also put that to the side for the moment. As we saw above, Sartre was happy to grant that we often arrive at putatively-determinate choice points, but the sense of self and commitments which set those determinate boundaries can themselves be traced to some previous, arbitrary, exercise of will. I am grateful to Ben Laurence and Gabriel Richardson Lear for discussion of this point.

instincts. But these, too, offer no guidance, for the very truth about our emotions is manifested in choice rather than something immediately introspectively available. Or at least this is how it is for the student:

> I can only estimate the strength of [some] affection if I have performed an action by which it is defined and ratified. But if I then appeal to this affection to justify my action, I find myself drawn into a vicious circle. [...] ...feeling is formed by the deeds that one does; therefore I cannot consult it as a guide to action. And that is to say that I can neither seek within myself for an authentic impulse to action, nor can I expect, from some ethic, formulae that will enable me to act.[186]

So Sartre concludes that his student knew all along that there was no answer. When he looked to his options, trying to see *himself* in one rather than the other, there was, in a deep sense, nothing to see:

> ...in coming to me, he knew what advice I should give him, and I had but one reply to make. You are free, therefore choose, that is to say, invent. No rule of general morality can show you what you ought to do: no signs are vouchsafed in this world. [...] For the decipherment of [whatever signs we think we see], however, [we bear] the entire responsibility.[187]

The student could not but see the choice as requiring the affirmation of a particular identity over another, yet finding nothing in himself which directed him toward one specification of his existing self rather than another, all he could do was fling himself forward into the void. Even so, the burden of responsibility could not be shirked for this ostensibly arbitrary choice; it would call out as a sign to be deciphered, to be fit, at the

---

[186] Sartre 1946.

[187] Sartre 1946.

moment of choice and then again after the fact, into a broader pattern of choices and affirmations which constitute the only sense of self we have in a normatively empty world.

Let us try to be concrete about the normative status of the choice. Prior to action, there was no fact of the matter about what the student ought to do. The polestar of existentialist normativity—the self we fashion through our action and self-reflection—underdetermined the choice. Indeed, it *must* have, given the arbitrariness that haunts the adoption of core principles that structure future choices. Before action, the choice was normatively underdetermined, in the shallow sense that the student's existing values offered incomplete guidance, and in the deeper sense that those values only existed as 'his own' due to arbitrary self-launching at an earlier date. Through the psychological inescapability of the classical perspective on action, the normative status of the choice shifts at the moment it is made: the option selected takes on the cast of an *affirmed* option, a cast which can only be redeemed by backwards-looking considerations capable of knitting the choice into a broader, richer, temporally-extended story of the person who made it. And after choice, the normative status of the option selected—and therefore of the choice as a whole—is either indeterminate until the moment of death, or can grow or diminish in concreteness as reflective retrospection demands.

One thing to notice about this multi-layered underdetermination is that even if the exercise of normative bootstrapping favored by the normative voluntarist were possible, it would eliminate only one of these layers: a new self would give new sense to the issue of what ought to be done, while deeper worries about the arbitrariness of the self prior to specification would remain untouched. Further, given the existentialist commitment to a

narratively-structured or pattern-based criterion of some option's being 'the thing to do,' there is an even deeper underdetermination that undermines *any* putative normative status. If it is true that any particular status is narrative-dependent, and narratives themselves are constantly shifting given the inexorable flow of the river of time, even these statuses—as arbitrary or non-arbitrary, sensible or nonsensical, part of a coherent whole or a random blip—are subject to revision and destabilization at any point.

Now, it is only this deepest level of instability that could, as a criterion of rational intelligibility, possibly counter the rational arbitrariness of self-launching choices like that of **kierkegaard's editor** which give us the sense of self which still underdetermines choices like that of **sartre's student**—which is of course why we are exploring it in the first place. And the narrative criteria which constitute this foundational level must (on a deep level at least) be up to the individual using them to structure her own self-understanding. As commonly understood, narrative criteria have to do with what makes for *a good story*, one which is not only causally ordered in a sensible way but involves a certain kind of part-whole relationship and temporal emotional structure: crisis, resolution, denouement; challenge, defeat, reprieve; a beautiful tragic ending; a surprisingly happy conclusion.[188]

For the existentialist, of course, narrative criteria are further restricted by the putatively inescapable classical perspective on choice. There needs to be a definite sense in which our choices track options that were 'the thing to do' in light of a broader, wider picture of the life as a whole; as *necessary* to overall story which is being affirmed. This claim exists uncomfortably alongside the broader heroic insistence on our boundless normative

[188] Lloyd 2008; Velleman 2006.

creativity, as already highlighted: on the one hand, "even the blunders of life [must] have their own meaning and value,"[189] yet meaning and value are themselves constrained by "rational thought," which involves "interpretation according to a scheme that we cannot throw off."[190] Or again, "we are unable ever to choose the worse," so that "what we choose is always the better," and the choices we make are inescapably "creative … of an image of man such as he believes he ought to be," colored by an implicitly affirmative cast.[191] As Nietzsche says, although

> there is no trick which enables us to turn a poor virtue into a rich and overflowing one … *we can reinterpret its poverty into a necessity* so that … we no longer sulk at fate on its account.[192]

So we find that two essentially independent criteria are intertwined. Each step along the way must be seen not only as 'the thing to have done' at that particular point (though we could not justly have affirmed it as such at the time), but also as *necessary* to the overall narrative arc of the life of which it is a part. The connection is forged by affirming the whole pattern of a life; what is necessary becomes 'good' in the sense that it was 'the thing to do' in the service of the greater shape of the life, even if normative materials and facts to justify such a verdict in the ordinary sense are thin on the ground for certain choices. But this connection is sensible—even required—by the existentialist dictum that our foundational commitments are themselves rationally arbitrary.

---

[189] Nietzsche, *Ecce Homo*, II, 9.

[190] Nietzsche, *Will to Power*, §522.

[191] Sartre 1946. Compare, later: "Man makes himself; he is not found ready-made; he makes himself by the choice of his morality, and he cannot but choose a morality, such is the pressure of circumstances upon him. We define man only in relation to his commitments…"

[192] Nietzsche, *The Gay Science*, §17, emphasis mine.

The heroic existentialist proposal, recall, involves nested layers of underdetermination, arbitrariness upon arbitrariness. The intelligibility of individual choices involves their fitting a satisfactory pattern relative to those which precede and follow them; the normative legitimacy of those patterns depends on whether their sequence as a whole, up to the current moment or to the end of a life, could be affirmed by the person who created them. Intelligibility is still comparative; what has been given up is the possibility of any contemporaneous criterion of comparison. The synchronous balance of reasons will, in a thin sense frequently, and in a deep sense *always*, underdetermine the normatively appropriate outcome.

Affirmative appraisal of a life's narrative arc is inherited by the individual choices which comprise it. *All* choices must be counted: what is up to us is the sense we make of them, as intelligible and (therefore) justified by the role they play in contributing to a greater whole. And we must see our choices as reflecting a conception of ourselves—perhaps all of humankind—as we believe we *ought* to be. So in the end there is no room in existentialist narrative-building for waywardness that is not in some way directly redeemed by an overall sense of life as it should be lived. To fit the formula of human greatness, we must be able to say of every choice, '*it was absolutely the thing to do, though I couldn't have known it at the time*.' And the way in which it was 'the thing to' must admit of this kind of explanation: '*I see it now—even began to do so in the moment of choice itself—as reflecting not only my authentic self, but what I affirm as right and good in myself, and judge to be right and good and others*.'

It turns out, then, that the heroic credo that our choices have only the sense that we give them places some very classical constraints on what sort of 'sense' a choice might have.

Not any story will do, only one which culminates in a sigh of self-satisfaction. (This is so whether we suppose that the culmination of the arc is the end of a life, or simply any future vantage point from which past choices can be seen to display a kind of a pattern or point.) Ultimately, it is a credo that demands both too much and too little. Although our task in life is to "paint our own portrait," the tools we have been given to make this sketch are a poor match for the job.[193] On the one hand, I will argue, the standards it imposes on narrative interpretation could scarcely be met without significant acts of self-deception; on the other, it is far from clear that these constraints on our subjectivity, limiting though they are, can justly be used to inform a new standard of *rationality* in choice.(

We return to the example of **sartre's student**, torn between a life caring for his ailing mother and one spent on the front lines of *La Résistance*, to illuminate these concerns. Let's begin with the concern about self-deception. According to the **heroic existentialist**, the student must choose without rational guidance, hurling himself forward into the void. Let us suppose he joins the Resistance, engaging in guerrilla warfare against the occupying Nazis. His choice of a new life blowing up bridges, derailing trains, and otherwise obstructing the German effort must be one he is able to affirm, at some suitable point in the future, as not just right for *him* but reflective of an ideal of humanity he is proud to inhabit. The society he joins will be radically different from the quiet life with his mother that he leaves behind, and his character will indubitably be shaped by the experiences he has there. It is the choices he makes in this new world, and the cast he gives those

[193] "In life, a man commits himself, draws his own portrait, and there is nothing but that portrait" (Sartre 1946).

subsequent choices, that can retrospectively render the comparative choiceworthiness of the past itself. But how?

There are many known psychological mechanisms which automatically favor this kind of thinking about our situation. Selective or false memory,[194] hindsight bias,[195] choice-supportive bias,[196] consistency bias, egocentric bias,[197] the endowment effect,[198] outcome bias,[199] post-purchase rationalization,[200] rosy retrospection,[201] and subjective validation[202] are just a few such tools. Together, they form a powerful nexus of means we have to make ourselves comfortable with the past—including choices undertaken with the feeling that they were underdetermined or even suboptimal, irrational, at the time. Choice-supportive bias, which sits at the heart of this nexus, has repeatedly been shown to be a dominant mode of selective memory and redescription of past choices. As the name suggests, it is a pattern of misremembrance which sees the option selected as more favored by rational considerations after the fact than it was deemed to be before the choice was made.[203]

---

[194] Dellarosa and Bourne 1984.

[195] Kahneman *et al.* 1982.

[196] Ross 1989; Mather *et al.* 2000.

[197] Kahneman *et al.* 1982.

[198] Tversky and Kahneman 1991.

[199] Taylor and Brown 1999.

[200] Tversky and Kahneman 1991; Taylor and Brown 1999.

[201] Ross 1989.

[202] Kahneman *et al.* 1982.

[203] Mather *et al.* 2000.

Has the existentialist, then, simply made a virtue out of this kind of self-deception, elevating a particularly insidious form of *mauvaise foi*? This is a common criticism of the historical existentialists, especially Sartre, but it would be too hasty a judgment of the heroic existentialist as I have sketched him here, whatever Sartre's actual views. In fact, the stern optimism of the view is meant to strip away all such weakness and self-deception. When we paint the portrait of the self, the only brush we have on hand is *will*, the sheer will of the manifestly underdetermined choice. The arc of our lives is inscribed by all and only the choices we make. At the end of the day, it is not clear what space remains for the palette of rich interpretation that is supposedly the stuff intelligibility is made of, no room for creative story-telling when all rivers must lead inexorably to the present moment with no tributary unaccounted for and no oxbow lakes left to the side. The agent who manages a degree of intelligibility for any of her choices—even if we presume that retrospection can occur prior to the end of life itself—must be a hero indeed.

It is no trouble, of course, if we find the existentialist promoting a near-unattainable standard for the sort of human excellence that can impose order and form onto the chaos of self-conscious experience. The worry is deeper: if we are bound by the actual pattern of our past experience and the interpretative constraints of the classical perspective on choice, we do not get to deem some choices more important than others; we are not able to name some choices or entire periods of our lives as deviations or mistakes. The task seems well-nigh impossible once we also factor in the heroic position that the intelligibility of choice combines elements of self-affirmation with self-transcendence: that we see our choices not just as authentically our own but also as reflecting our judgments about how people, generally, ought to be.

I have been interpreting this view—classical indeed—as pointing us to the importance of publicity, of in-principle shareability, for the grounds of choice. Our choices cannot have been made on the balance of reasons, but we make sense of them by imputing to them all the important characteristics of this form of justification. That is why it is not enough to rest ultimate arbitration of the affirmability of a life in any single perspective; we must build into it the notion that the affirmation takes the form of seeing the choices as *redeemed* by their end product, the self and the story (such as it is) that they tell. But this, too, speaks against the possibility of any richer notion of self-mythologizing as the arbiter of consistency in choice. All we have is the pattern of our actual choices, strung like beads behind us, nothing knitting them together but a willful affirmation of their bare existence.

In a way, this concern that the heroic picture demands too much is just the flip side of the more familiar worry that it demands too little. Although retrospection is constrained in severe and perhaps impossible ways, when this process settles the question of which choices were intelligible and meaningful and to what degree, it takes the form, not of reflection, but of *affirmation*. We do not find, at the end of life, that our choices form a sensible pattern, for there is no criterion of 'sensibility' outside of our own subjectivity. We may weave our own standards, but these are groundless—not held firm by something bigger and stronger than ourselves, like God or human nature—but only by the choice, *which must itself be underdetermined*, to say 'yes' to them instead of 'no.' It is sheer will all the way down.

And so the stern optimism of the existentialist position is either too stern or too optimistic. Sternness offers structure but blots out the creativity retrospection seems to

require; optimism about the will's power to explain the mechanism of retrospective recalibration but bottoms out in an affirmation which stands in stark contrast to the mid-level insistence on a classical form of choice.

A deeper worry looms just behind these others, and unites them. There seems to be something fishy, temporally, about the entire picture. In a deep sense, there is simply no normative fact of the matter about what the balance of reasons favors prior to the student's choice. The moment of choice itself does nothing to alter this balance; the act of will that propels him out of deliberation and into the future is sheer, not normatively creative. And on the heroic model, the normative status of his choice *must* remain open to future developments. It is only retrospection that establishes the intelligibility of the choice made, and thereby settles the matter of what there was, in fact, most reason to have done at the time. But the kind of intelligibility the choice gets is that of '*worth doing*' because of its role in an overall pattern of choice—and a sense of normative identity that emerges from this pattern—which can be affirmed from a suitable distance in the future. It is manifestly not an account of *what made the choice worth doing at the time in the eyes of the person making it*, nor could it be.

But then there is something distinctly odd about saying that, in choosing, we choose in a way that both creates and affirms an image of humanity as we believe it should be. Consider again the poor **student**: had he a conception of 'humanity as it should be' capable of answering the question of what to do, he'd never have found himself consulting his philosophy teacher for help. Since his existing conception failed him—on the surface level we have described and on the deep level that even his existing ideals must themselves have

once been arbitrarily chosen—his choice will express nothing but the necessity of making a blind leap, however much he may later be grateful for having done so, and indeed find it impossible to do anything but, from the satisfied perspective of the person he has become, affirm that choice.

Put another way, the deeper worry in the existentialist account is that its narrative tools seem to require that choice be expressive of a self in process—a self being forged by the choices made, rather than guiding them by shaping the balance of reasons prior to deliberation itself. And yet if the self is truly in process, never 'being' but always 'becoming,' there is nothing for the choice to *express* at the time it is made. The ontology of the self that shimmers behind the existentialist view is disturbingly thin: all will and choice, with (as we saw) little space for narrative flexibility and apparently no role left for standard rational processes to play in the unfolding of our lives. This is different than worrying that the existential achievement of self-intelligibility requires bad faith; it goes right down to the bare conceptual possibility of such self-understanding. Our choices do not reflect the balance of reasons as we saw it at the time of choice—as it *was* at the time of choice—but rather a conception of ourselves which necessarily blossomed and grew from them only after the fact. In a deep sense, there could be nothing for those choices *to* express.

The fact that we seem to *need* to see our choices as self-expressive cannot be treated as simply self-ratifying—a notion to which we shall return in the next section. This need, and the impossibility of meeting it honestly, might simply be another element of the absurdity of human experience. The heroic existentialist fixation on classical constraints like comparative choiceworthiness includes only the inescapability of a certain kind of

perspective on choice: the classical model characterizes neither the mechanism of conscious deliberation nor the form of affirmation which, at the end of the day, stands in for a classical conception of rational ordering in choice. The *constructivist* account of the normative authority of self-affirmation, by contrast, makes far more prominent use of these classical constraints. As with the heroic position, this affirmation must, at the end of the day, be a shout into the void rather than a recognition of anything binding outside of the self. But the rational constraints on our deliberative perspective do more direct work in establishing both the legitimacy of our subjectivity and the rationality of our agency. Reason is not left quite so far behind.

4.4 Existentialist constructivism

Constructivism, as a metaethical position, has a variety of forms on the contemporary scene. The one of interest to us is the existentialist-tinged form espoused by theorists like Christine Korsgaard, and it is of interest to us not because of its metanormative claims but because of the theories of agency, identity, and rationality that underwrite them. The existentialist strand of the constructivist project knits together key elements of the existentialist understanding of human action—the psychological inescapability of a certain kind of perspective on choice, the normative bankruptcy of the world as we find it—but envisages a different role for will in underdetermined choice. Like the heroic existentialist, the constructivist sees rational underdetermination as a pervasive condition of the human experience. It is not a peculiar, quasi-mathematical phenomenon limited to certain contexts, as the hybrid voluntarists would have it; but a structural feature of the perspective we inhabit—indeed, a condition on the very possibility of self-directed choice.

Constructivists in contemporary ethics and action theory make some notion of reflective endorsement or engaged volition central to their account of the normativity of practical reason itself.[204] On their view of decision-making, most or all human choices are underdetermined until some exercise of the will steps in to settle the question of what there is most reason to do: our choices, shaped by our inescapable practical perspective, force the creation of the normative entities thought to structure and guide them. Reasons, on this view, name the solution to a practical problem—*what should I do?*—but they are materials we fashion for the sake of answering this practical problem, not ones we discover as already normatively relevant to the choice we face.[205]

The motivation for constructivism is not typically the rending of the fabric of the old world that so thrilled and horrified the existentialists. Life in the normative abyss is these days more often explored on the basis of a scientifically-responsible philosophical naturalism: a healthy suspicion of brute normative facts or entities mixed in amongst the natural facts of the modern scientific world-view. The claim is that normative force itself is inexplicable without some account of how we can bind *ourselves*—a claim in many ways foreshadowed by Cudworth's argument of the last chapter—without which the normative entities of the world would have neither authority nor motivational grip. Only a law we give ourselves can truly bind; only the principles we choose for ourselves can truly motivate.[206]

---

[204] Contemporary constructivists include Wallace 2006; Street (ms), 2008, 2010; among others.

[205] As Korsgaard often puts it, reason is the name for a kind of "reflective success" (Korsgaard 1996a, e.g. §§3.2.1–3, 3.3.2, pp. 92–98, 102.

[206] Wallace 1999; Korsgaard 1996b.

A view of life in the void as a uniquely human struggle, however, is a more recognizably existentialist inheritance, about which Korsgaard is explicit,[207] and on which our exploration will focus. On her view, self-conscious rational agents strive to make choices for reasons that can be in a certain sense *shared*, ones that involve, as Sartre says, implicit endorsement of a particular picture of human life. The necessitation we feel in choosing to act carefully and well is inescapable, and ownership of our choices would seem to be an essential precondition for responsibility, moral or otherwise:

> For my movement to be my action, for it to be expressive of *myself* in the way that an action must be, it must result from my entire nature working as an integrated whole.[208]

Yet because of the world we find ourselves in—as structured by our own subjectivity as self-conscious, reflective creatures, by our inescapable first-personal perspective—the sought-after connection between self and choice is unavailable.[209] For the heroic existentialist, a connection between the world (or a balance of reasons determined by it) the choice itself was what was sought, a connection which the empty world could not

---

[207] In her reply to Bernard Williams, Korsgaard affirms Thomas Nagel's characterization of her view as "rather existentialist" (though denies that this makes it, as he claims, "rather unKantian" as a result). Original mention in Nagel 1996, p. 203; affirmation in Korsgaard 1996a, "Reply [to Bernard Williams]," p. 237.

[208] Korsgaard 2009, §1.4.1, p. 19; *cf.* §1.4.2, p. 19: "We [human beings] are self-conscious in a particular way: we are conscious of the grounds on which we act, and therefore are in control of them. [...] When you deliberately decide what sorts of effects you will bring about in the world, you are also deliberately deciding what sort of a cause you will be. And that means you are deciding who you are. So we are each faced with the task of constructing a peculiar, individual kind of identity—personal or practical identity... *It is this sort of identity that makes sense of our practice of holding people responsible, and of the kinds of personal relationships that depend on that practice*" (emphasis mine).

[209] Korsgaard 2009, §1.4.3, pp. 19–20.

furnish. For the constructivist, this connection must hold between the self (and balance of reasons) and choice. But it cannot:

> You might suppose that [intelligible, attributable choice] requires that an action be the effect or result of a *prior* unity in the agent, an integrity already achieved. You first achieve the sort of psychic unity or integrity that makes you the master of your own movements, that is, that makes some of your movements attributable to you as *yours*, and then the choices that lead to your actions express the unified selfhood you have already achieved. But … this cannot be how it works. […] in the relevant sense there is no *you* prior to your choices and actions, because your identity is in a quite literal way *constituted* by your choices and actions.[210]

Just like the old heroes, the constructivist looks to the natural materials of normative justification and finds that she has been furnished with none prior to choice.

Yet unlike the old heroes, the constructivist sees the activity of the will as *creative* rather than *sheer*. The role of the will in underdetermined contexts—that is to say, the role of the will in all choices—is not to fling us forward into the void, leaving us to make sense of that action retrospectively, if at all. Rather, as the quotation above makes clear, its role is to move us forward *on the basis of* normative materials forged by the activity (and in the moment) of choice itself. Like the normative voluntarist, the constructivist sees reflective choice and practical identity as mutually-adaptable pieces of an agent's deliberate movement through space over time.[211]

---

[210] Korsgaard 2009, §1.4.2 (p. 19).

[211] "It is as the possessor of personal or practical identity that you are the author of your actions, and responsible for them. And yet at the same time it is in choosing your actions that you create that identity. What this means is that you constitute yourself as the author of your actions in the very act of choosing them." (Korsgaard 2009, §1.4.3 [pp. 19–20]), *cf.*, immediately below, her pronouncement that "judgments of responsibility don't really make sense unless people create themselves" (*ibid.*).

And just like that of the normative voluntarist, the constructivist proposal sounds mystical —if not straightforwardly paradoxical. How can you create yourself, as the true author of your choice, unless you are in a certain sense already there? The earlier existentialist picture had an air of mystery—what new standard of normative intelligibility could work backwards in time? But this new one can sound even more absurd: what minor miracles of simultaneous creation and justification must we be able to perform if the constructivist is right?

The dilemma of choice for the agent is twofold, on the constructivist picture. First, we must make a choice in the absence of already-existing normative guidance. All we have are the tools we can make ourselves—but we can only make those tools by completing the choice we seem to need the tools to make. The only source of normative guidance available to us—normative principles or reasons generated by the self, in its ongoing struggle to function as a unified whole—is one we must forge *in the very act of choosing*. What can this mean?

According to Korsgaard, every choice is at the same time an attempt at agential unification and self-understanding. Any discrete choice is inseparable from this continuous effort to be a single unified agent, rather than a scattered set of impulses and half-abandoned ideals.[212] Understanding the process of *becoming* who one is as essentially directed at a state of *being* is something that Nietzsche, Sartre, and Korsgaard alike aim to unsettle. Recall the passage cited earlier, in which Sartre analogizes life to art, where he seems to insist both that the painting *becomes what it is at the moment of creation* and that *it can be made sense of only*

[212] Korsgaard 2009, §I.I.5 (p. 7).

when complete.[213] Nietzsche's oft-quoted dictum that one must 'become what one is'[214] destabilizes the relation between being and becoming in just the same way:

> Becoming must be explained without recourse to final intentions… Becoming does does not aim at a final state, does not flow into 'being.'[215]

Korsgaard's proposal adapts Aristotelian teleology to the realm of human agency. Being a self—having an identity—is not a state but a process of continual becoming. Just as, for Aristotle, to be a living thing is to be engaged in a continuous process of nutrition and self-perpetuation, for Korsgaard, to be a rational agent is to be engaged in a continuous process of self-constitution, of affirming and re-affirming the collection of practical identities which comprise our self-conception, through the act of deliberate choice. In acting, we constitute ourselves as the authors of our choices; when we make choices dictated by the policies of our practical identities, we affirm our commitment to them, and

---

[213] "No one can tell what the painting of tomorrow will be like; one cannot judge a painting until it is done. […] we understand very well that the composition became what it is at the time when [it was being painted]" (Sartre 1946).

[214] "How one becomes what one is" is the subtitle of *Ecce Homo*—ironically Nietzsche's final published work—but the theme recurs often in earlier work (e.g., *The Gay Science* §§270, 335; *Thus Spake Zarathustra*, "The Honey Sacrifice," IV, 1). For discussion see Nehamas 1985, ch. 6, esp. pp. 171–73. I borrow the observation about irony from Nehamas (*ibid.*).

[215] Nietzsche, *The Will to Power*, §708; *cf.* §552: "'Truth' is … not something there, that might be found or discovered—but something that must be created and that gives a name to a process, or rather to a *will* to overcome that has in itself no end—introducing truth, as a processes *in infinitum*, an active determining—not a becoming conscious of something that is in itself firm and determined"; compare also his self-stated aim "to transform the belief 'it *is* thus and thus' into the will 'it *shall become* thus and thus'" (§593).

ought further to aim at unifying and integrating the identities with one another, and with our core practical identity: that of rational reflective agent, of 'human.'[216]

For the constructivist, as for any other radical, constrained subjectivity is key to making sense of our choices, and the lives they comprise. Our authority in this matter is in part explained by there being no other source for meaning available: our acts have only the sense that we give them, for there is no other power on heaven or earth to grant it them. The constructivist project requires that even if such powers *were* regnant, they could not justly reign over *us*, because the reflective structure of human subjectivity forbids normative constraint by any force external to the rational deliberator herself. Put another way, the constructivist insists more on the power of the will than the emptiness of the world, and views the will as directly creating the normative materials we use to make sense of choice.

The will has this creative power because of what it is to make a self-conscious choice. Whenever we choose deliberately—whenever we undertake action after some pause, however brief or even invisible to the conscious mind—we must see ourselves as choosing *on the basis of* some consideration or other. The constructivist holds that deliberation, by its nature, involves automatic suspension of whatever motivational pull the various considerations at play have exercised on us up to that point. There is a distance between the agent's will and any consideration that offers itself up for inspection (say: '*join the Resistance, it would be the noble thing to do*'). Its normative pull is not automatic, and its motivational pull ceases to be automatic at the moment we pause to call it into question. It

[216] Korsgaard 2009, §§1–2; see esp. §2.4.1. The Aristotelian analogue appears in earlier work as well, see Korsgaard 1996a, §4.3.9, esp. p. 152.

can exercise the requisite normative and motivational pull only after we endorse it as something that *really is* a reason to do the thing in question.[217]

Moving from deliberation to action on the basis of some consideration (or bundle of considerations) you've named as capable of justifying that course of action is what shifts its normative status from inert to rationally-authoritative in the traditional sense. Those considerations serve to explain and justify the choice made, to the person who made it and to others around her. There is, strictly speaking, no fact of the matter about what the balance of reasons favors prior to the termination of deliberation and the initiation of action, because there is no balance of reasons until the individual deliberator has decided which considerations will play the role of 'reason' in the full story of how and why she acted the way that she did.

The considerations that offer themselves up as candidates for the role of 'reason' for any given choice are shaped by the interface of the world with our existing complex of practical identities and normative ideals. If I have long considered nobility an important moral virtue, one that either is or ought to be at the center of my life choices, considerations of the form '*x* would be noble' are likely to be both salient and (when I grant them this authority) weighty in my deliberation. There are no reasons for action which are not in some way connected to one of my identities, including my animal and human nature.[218]

---

[217] Korsgaard 1996a, §3.3.1, p. 100.

[218] Take **aristotle's dog**, for example, torn between food and water by equally-forceful hunger and thirst. On the constructivist view, the animal only has a *reason* to eat if he endorses the nature which makes hunger salient to him (presenting the hunger as a motive to pursue food) and justifies it as a motive for action (since this animal nature involves self-maintenance as a constitutive norm). The reason does not spring, full-formed, from either his desire or the nutritive value of the food.

These constraints on my choice are only as binding as those identities are essential to my self. If I find I cannot throw off a particular identity, then the considerations it deems weighty cannot but be weighty for me in choice; when I choose in ways that violate the ideals of a particular identity, I fail to act with integrity; making choices at odds with my ideals erodes my commitment to the self-conception that names them worthy of pursuit in the first place. The reasons for which we act *express* our identity.[219]

Like Nozick, the constructivist notes something interesting about personal identity in connection with reasons for action: that our self-consciousness makes identity a reflexive notion. One's self-conception is not just a conception of the self, but the conception that a self has *of itself*. The reflexivity of personal identity was crucial to understanding the "self-subsuming" nature of choices made on the basis of personal reasons, even if we ultimately found the appeal to self-subsumption to be a dead end. The constructivist account invokes something similar in its theory of the normativity of personal identity: our identities are only binding on us when we affirm or endorse them, so normative force is still located within the self rather than outside it, but because our identities are not easily surrendered, and some notion of identity is required for coherent choice in the first place, they have a kind of motivational and (in the case of human identity) justificatory grip on action.

And so the constructivist grounds rationality in agency rather than agency in rationality, a complete reversal of the classical rationalist view and an embrace of the truly voluntaristic half of the hybrid voluntarist project. Constructivism filters this account of the balance of reasons through a notion of self-conscious identity, but the constraints the interpretation

---

[219] "Your reasons express your identity, your nature; your obligations spring from what that identity forbids" (Korsgaard 1996a), §3.3.1, p. 101.

of choice are not the diachronic, retrospective, narrative, pattern-based constraints of the heroic model. These constraints are synchronic; they spring from the self that makes the choice at the moment she makes it. Although the self is essentially fluid—always becoming and never being, made and unmade with every choice—the reasons for which it acts are self-referential. Our choices are at once an expression and an affirmation of the identity that makes them sensible to us.[220]

4.5 Problems for the view

There is a very old concern about grounding the authority or legitimacy of normative concepts in the will of the individual they are meant to bind or govern. This concern is often raised against constructivism's metanormative claims. I will suggest that a version of this problem applies to the existentialist-tinged constructivism under examination—as a challenge *not* to its theory of practical normativity, but rather to the view of rational agency on which the normative theory depends.

As Kant puts it, introducing the antinomy of self-directed obligation, the basic concern is this:

> If the 'I' *that imposes obligation* is taken in the same sense as the 'I' *that is put under obligation*, a duty to oneself is a self-contradictory concept. …the one imposing obligation (*actor obligationis*) could always release the one put under obligation

[220] For example: of my many nested practical identities, here are two: I am a sister and a daughter. These familial roles places considerations that might not even be salient to other people ('tomorrow is Teresa Heuer's birthday,' 'Amy Heuer needs help finding an apartment') well inside my deliberative field, and suggests that they count in favor of certain actions and plans and against others. I might find it psychologically difficult to renounce certain identities—family ties are a prime example—but on this view I may disown them by rejecting the constraints they aim to place on my behavior. I am bound only if I allow myself to be bound.

> (*subjectum obligationis*) from the obligation (*terminus obligationis*), so that (if both are one and the same subject) he would not be bound at all to a duty he lays upon himself. This involves a contradiction. [6:417]

For the existentialist constructivist, self-binding flows through the concept of personal identity, rather than the direct and self-conscious issue of laws. But this seems only to heighten the basic concern. The agent whose identity is meant to be taken as authoritative only has that identity so long as she affirms it as truly her own. Its normative authority resides in her identification with it, but this identification is an exercise of volition rather than observation, perception, recognition, or some other process that can go right or wrong. We cannot be in error about which identities are truly ours, since it is only our will that makes any identity ours in the first place. Moreover, since it is only our will that adheres us to one self-conception rather than another, it seems that any self-conception will do: what then is the sense of 'authenticity' or 'ownership' that roots our choices in a normative self-conception if it can evaporate at a moment?

These two concerns are cousins of the worry that if our reasons are willed rather than world-given, we can will ourselves into identities (and thereby give ourselves reasons) that are truly dreadful.[221] Korsgaard, a Kantian as well as an existentialist, takes herself to have an adequate reply to this concern: commitment to our human identity involves, ultimately, commitment to the cause of humanity itself. One cannot coherently will to be an evildoer (pirate, mafioso, dictator), *ergo* one cannot give oneself reasons for doing evil things. Since commitment to our human identity is a prerequisite for having any other identity, and

---

[221] G.A. Cohen's comments in Korsgaard 1996a are the first and in some ways still the fullest description of this concern, but it has been echoed by many since. See, e.g., Cohon 2000, Fitzpatrick 2005, Wallace 2006.

therefore any other reasons at all, we all have reason to obey the moral law, even though the law does not spring from any source outside of our structured subjectivity.[222]

But this response does not help with the worry about the normative authority of our contingent practical identities, which must be the source and guarantor of reasons in any practical context to which moral considerations do not apply or themselves underdetermine the appropriate outcome. If these contingent identities can be adopted or abandoned at will—literally, *through an exercise of volition*, no matter how well-considered or infrequent—then in the end it is still only my will that makes certain considerations into reasons for action. The situation that Kant identified—with nothing barring the *actor obligationis* from releasing the *subjectum obligationis* from the constraints of her extant practical identity and thereby altering the balance of reasons in any direction—stands. This is the sense in which, for the constructivist, literally all of our choices are rationally underdetermined until we make a choice under a certain description of ourselves. But with nothing constraining which contingent description of ourselves we opt to endorse, we face yet another and deeper level of underdetermination. And unlike the heroic existentialist, who sees our adopted self-narrative as constrained by the pattern of our actual past choices, the constructivist has no such theory to delimit the appropriate bounds of intelligibility-conferring, criterion-constituting choice.

Let me put it another way. In the classical model of choice, our identities shape our reasons, and our reasons guide choice. This sounds very much like the pattern suggested by the constructivist—and yet our identities are fluid rather than fixed, and so seem ill-suited

---

[222] This is putting it all quite loosely—for more see Korsgaard 2009, *supra n.* 75.

to play the role of fixing the balance of reasons that supposedly constrains and directs choice-making. For the constructivist, our identities are not only radically fluid but are in a way only themselves fixed by action. Since we humans are self-conscious, ownership of the choices we make, self-expression, and self-constitution are all rolled up into one process, that of deliberate choice.[223] Yet the materials of deliberation we have on hand prior to choice—our identities and the considerations they suggest to us as having normative weight—can only play the role in deliberation that we say they do. In choosing, we endorse an identity we do not properly own until it is endorsed, yet this identification persists only until the next time we face the necessity of action, when any identity but one (our human identity) can once again be called into question. And since it is our plight always to be faced with the necessity of action, the self is no more than an ephemeron, an illusion.

What would seem to be required is some constraint on the adoption or dissolution of our contingent identities: that they be more or less consistent across time, that our selves are more or less continuous with what comes before or after, that their alteration proceed piecemeal rather than wholesale, or that their development follow some antecedently- or separately-endorsed pattern of maturation. But this is directly precluded by the existentialist view. Our exercises of will may be constrained formally—in order to count as exercises of agency at all, according to Korsgaard, they must succeed in making us into the autonomous and efficacious cause of our own action—but not in any substantive way. And so they cannot be guided or delimited by the content of past choices unless we

[223] Korsgaard 2009, §I.4.2, p. 19.

ourselves will that they be so guided. Parts of our extant identity may function as 'fixed' points only if we treat them as untouchable—but this is itself a choice, not a capitulation.

It is a choice, moreover, that we are in a certain sense *wrong* to make, for it misrepresents the nature of the acting self. We indubitably face strong psychological pressure to treat certain identities as authoritative in action because we feel them to be so deeply our own that we could not possibly disown them; identities like 'older sister,' 'Christian,' or 'Mexican–American' spring to mind. Indeed, some critics of constructivism have argued that our identification with self-conceptions takes exactly this form: some deep identities *automatically* govern choice, without any exercise of the will needed to reify the considerations they offer as potential reasons.[224] And, as we saw in §4.4, we face considerable psychological pressure to treat our own choices as justified and our own identities as unified across time. But if the existentialist is right, then capitulation to these psychological pressures will be nothing more than another exercise of *mauvaise foi*, an unacceptable slide from the 'is' of psychological comfort to the 'ought' of normative legitimacy. Nothing has an automatic grip on us but the bare constraints of willing themselves.

Can these simple constraints demand that we treat our self-conceptions as basically unassailable, at some times, or in some places, for the sake of shaping the normative terrain we face in choice? It seems that I would need to treat *some* constellation of more determinate identities as unassailable in the moment of choice in order for that self-

[224] Cohon 2000 proposes this directly; recent discussion bringing together Harry Frankfurt's work on 'volitional necessities' and Korsgaard's constructivism has a similar theme. See, e.g., Watson 2002.

conception to be expressed therein, for there to be a balance of reasons that does not strictly speaking precede but at least accompanies the decision to act. But this again is illegitimate, according the radical credo that the self is a process rather than a product: without some further normative warrant standing behind this formal requirement that we treat the self as a fixed state rather than a fluid potentiality, it remains an error rather than a source of legitimate authority.

Now, the goal of integrity—of the unification of the diverse factions of impulses and ideals that comprise our conscious experience of ourselves—is one we also implicitly hold in acting, Korsgaard argues.[225] But this is *synchronic*, not *diachronic* integrity. We aim at unity in action in the sense that we want our actions to reflect the activity of our whole selves, not that our activity across time reflects a consistent substratum of unchanging (or only gradually-shifting) identity. And while we are, on the Kantian view at least, constrained in our choices by commitment to a conception of ourselves as human, the decision that selects which of the various instances of human value or human excellence (ballet dancer, scientific researcher, mother, manager) we choose to make our own cannot be constrained in the same way. If we adopt the identity of a dancer, or reaffirm an existing identity as a mother, perhaps we cannot but aim to be *good* instances of those kinds—and so are constrained in choice by what count as good-making features of various actions relative to those identities—but there is nothing that makes it more normatively appropriate to

[225] Korsgaard 2009, §4.4.

commit myself to one strand of the general human story (say, the ballet strand) than to any other, outside of my own will.[226]

This should concern us for the same reason that we were troubled by the sheer will at the foundation of the heroic existentialist view. For it seems that the constructivist has either engaged in the same suspect slide from psychological inescapability to normative authority, from the 'is' of what our choices must seem like to us to the 'ought' of genuine normative constraint, or has failed to offer an account of the rational content of anything other than morally-determinate choices. I have not explored the argument for the rational authority of moral norms on this view, and I take no stand on it now, for even if the constructivist is right that the authority of morality comes from an exercise of inescapable volitional commitment to our own human identity in a way that carries with it unavoidable respect for the well-being of others, this answers only the question of how morally-constrained choice-making can still be agent-guided. The question of how our contingent identities are forged by choice remains mysterious—and the rational content of underdetermined choices, the content which renders them intelligible to ourselves and to others, out of reach. The constructivist leaves reason behind in a different way than the old heroes, did, a movement away from normative constraint that is both subtler and more complex. But at the end, it amounts to the same thing: a compelling account of human agency that leaves

[226] Korsgaard 2009, §10.1.7, see esp. p. 212: "…in being the author of your own actions, you are also a co-author of the human story, our collective, public, story. As a person, who has to make himself into a particular person, you get to write one of the parts in the general human story, to create the role of one of the people you think it would be good to have in that story. And then—at least if you manage to maintain your integrity—you get to play the part."

little to no room for the constraints of intelligibility-conferring *rationality* to get a grip without engaging in a new form of *mauvaise foi* about the self.

4.6 Conclusion

The existentialist conception of agency in the face of the disintegration of rational guidance offers what is in many ways a lovely model of the role of choice in human life: optimistic rather than despairing, celebrating the creativity and spontaneity of the human experience, with full attention to the aspects of self-consciousness and reflexivity that give life to the quest to understand underdetermination in the first place. Yet this account, as we have seen, mirrors the vices of the classical model of choice. The dogmatic rationalist derives a model of agency from the comparativist theory of rationality; the radical voluntarist grounds a theory of rational intelligibility in a theory of free agency. And whereas the classicists left no room in their theory of rational intelligibility for a conception of what matters in agency, the existentialist move to ground the intelligibility of choice in a model of agency leaves little or no space for the intelligibility-conferring power of reasons to do its work. The dogmatic rationalist insists that there are no underdetermined choices; the heroic existentialist that there are no determined ones. Little wonder, then, that their accounts of human agency are so dissimilar.

The celebration of rational arbitrariness is not without appeal. Perhaps there is something good in the bare exercise of caprice, of simply willing in the absence of rational guidance. Something like this idea animates the exhortation to throw off the shackles of rational justification we find in Dostoyevsky's underground man, for whom such choices preserve

individuality and personality and prove him to be "a man and not a piano-key."[227] We find the same thought in the triumphant conclusion of Francois Camoin's short story "American Literature," which in its denouement praises the protagonist's whimsical choice as marking the first time in his life he acted "for no reason at all."[228] Perhaps we tap into something special, and specially-human, when we launch ourselves forward by will alone.

There might also be instrumental value in living a life shot through with little moments of arbitrary, random choosing—it might perhaps enable a special kind of psychological flexibility, or have the benefit of exposing us to more and varied terrain in the world than were we to stall in the face of underdetermined choice. Something like this idea is implicit in Elijah Millgram's model of practical agency as depending on practical induction. Experience provides the material on which this induction is performed; we learn from our varied experiences what matters and what does not, and we become better and more unified agents as a result.[229] Since retaining some degree of vagueness or underspecification in self-governing rational policies will enable us to more easily adapt to new circumstances, and change is a basic feature of our practical environment here on earth, we have reason to retain policies which will themselves at least sometimes underdetermine the normatively-

---

[227] Dostoyevsky 1864, pp. 21–22: "this caprice of ours, may be in reality, gentlemen, more advantageous for us than anything else on earth, especially in certain cases… for in any circumstances it preserves for us what is most precious and most important—that is, our personality, our individuality. […] I believe in it, I answer for it, for the whole work of man really seems to consist in nothing but proving to himself every minute that he is a man and not a piano-key!"

[228] Camoin 1992. I am grateful to Elijah Millgram for the reference.

[229] Millgram 1999.

appropriate outcome. And so perhaps arbitrary choices might be central to our self-constitution as practically unified and yet adaptable human agents.

Yet rational arbitrariness cannot stand at the heart of what matters in human life, any more than external rational constraint can comprise its core. In the next and final chapter, I outline two routes out of this bind, but ultimately suggest that the only way to leave the stalemate behind is to abandon commitment to the comparativist model of rationality which has given life to all the forms of classicism and voluntarism surveyed so far.

5 Conclusion

To be a classical rationalist is to be a comparativist about rationality and a rationalist about agency. That is, rational choice requires guidance by a comparatively-rendered univocal balance of reasons, and rational choice-making is the paradigmatic exercise of human agency. On the classical view, comparative rational guidance exhausts the domain of practical rationality: where reasons underdetermine the outcome, we are robbed of the only materials that make choice intelligible. This dissertation has traced the slow decay of the classical picture, one half-life carrying us to hybrid voluntarism, the next to radical voluntarism, as the second tenet of the classical view—rationalism about agency—was gradually abandoned.

We left classical rationalism for voluntarist terrain on a hunch that will-based theories might offer a compelling answer to the problem of the missing agent. At every stage, though, the need for rational comparison as a source of normative intelligibility re-asserted

itself. The **dogmatic rationalists** could not account for ordinary acts of human foolishness, from akrasia all the way to mortal sin; more to the point, their view left no space for freely-willed, -owned, or -authored action. The hydraulic process by which reasons determined our choices precluded meaningful agential guidance. The **hybrid voluntarist** move was to reject hydraulicism in *just enough* cases to slot in some agency, while retaining a basically hydraulic description of the role of reasons in deliberate choice. This proved insufficient: when the **sheer voluntarists** tried to locate the agent in exercises of sheer will, they enabled arbitrary forward movement, not self-expression—little more than an unowned impulse. The **normative voluntarists** tried to implicate these impulses in sudden shifts of personal identity, reclaiming agential ownership, but became tangled in the partly-hydraulic model of choice and the conception of reasons which the model presupposes.

The **radical voluntarists** abandoned hydraulicism altogether: since the world never offers a balance of reasons capable of tipping us directly into choice, the role of the will is not to pick up the slack when reasons run out, but to launch us forward in the absence of rational guidance. Intelligibility is instead tied to an affirmed self-conception: our choices are justified by endorsement of the identity they express. For the **heroic existentialist** this is a diachronic process whereby the entire pattern of a life is affirmed; for the **existentialist constructivist** it is a synchronic process in which a unified self is affirmed as worthy of expression in the action undertaken. Yet we discovered at the heart of the radical project an uncomfortable form of *mauvaise foi*. For the heroic existentialist, the narrative affirmed involves misperceiving the balance of reasons as being a certain way at a time when, by definition, there was no balance. For the constructivist, affirmation involves seeing the self

as genuinely inhabited or necessarily binding in a way that we must know that it cannot be. And so no point along the spectrum from classical rationalism to radical voluntarism offered safe harbor: the cost of living in each territory on the conceptual map was too high to bear.

Are we then condemned to plant a flag in classical or radical terrain, making what peace we can with the impoverished model of agency or bad faith about rationality to which we are thereby committed? In this chapter I argue that we are not—but that the search for firmer ground must ultimately abandon the the very model of rational intelligibility which was our starting ground, making further exploration a task of monumental scale and uncertain scope.

The argument for this conclusion comes in two stages, corresponding to two key assumptions that set the boundaries of our original exploration.

The first assumption is *methodological*, and has structured the entire investigation. In each chapter we sought 'the missing agent' in action by examining whether the normative mechanics of the theory in view left anything like an agent-shaped hole: no hole, no agent, and so no acceptable theory of agency. In the first part of this chapter, we turn this methodological assumption on its head—beginning with a somewhat thicker notion of what it is to be an *agent* (in particular, what it is to be diachronically self-conscious) and seeing whether we can't find a model of rational choice-making that does justice to this view of personal identity, rather than trying to slot agency into an extant model of choice.

The second assumption is *conceptual*, and concerns the essentially comparative function of reasons (or the essentially comparative nature of practical rationality). The comparativist

view of rational choice is, as we have noted, the one half of the classical model held fixed so far—every theory here cleaved to this notion of rational choice, even as their notions of intelligibility in willed choice diverged. In the second part of this chapter, we ask whether the best way to find livable ground isn't to shed the comparativist standard entirely, and move to a more flexible notion of rational justification that allows other ways that reasons might contribute to the intelligibility of choice than by contributing their weight to a single rational scale.

We revisit these assumptions to see whether they haven't artificially constrained our view of the theoretical landscape. If, as I argue, they have not, our map's borders are the more firmly fixed, and departure from them all the more daunting. We conclude with an account of what this departure must involve, for those unwilling to take up residence in already-explored terrain.

5.1 Assumption one: methodological

It's no accident that the voluntarists found recourse to personal identity in their attempt to build theoretical support for the the phenomenologically obvious richness of agency. But turning to the 'self' in the quest for the 'agent' seemed to lead us over and again into paradox and confusion. The conjecture we test in this section is that starting from a notion of personal identity that incorporates the insights about temporality and self-consciousness highlighted by the voluntarists in §§3–4 might be the key that unlocks some habitable ground contiguous with, but distinct from, the proposals already explored.

To begin, consider another criticism of the classical model: that its more dogmatic instantiations have no obvious way to explain the *development* of personal identity over time. None of us are born full-formed; many of the traits that make us who we are—and which give rise to the sorts of identity-based reasons that can tip the rational balance—develop over time, through childhood into adulthood and senescence. Not only is self-development diachronic, sometimes it is even (it seems) *deliberate*: an existing self may consciously act in ways that bring about its own transformation. Shifts in character of this kind are often colloquially labeled 'exercises of will'—the smoker who quits, the liar who gives up her mendacious ways, the wallflower who becomes a confident leader—the thought perhaps being that reasons given by an existing identity cannot justify movement to another. It is the will, strung along by some guiding ideal, that presses us toward a better future.

Now consider the dogmatic dictum that one acts as one has most reason to do or does not act at all. Let us suppose, as I have been doing all along, that our reasons may come from our personal identities as well as from other sources: our dispositions, preferences, and overall character make some things count as reasons for us in the first place, considerations which might not count that way to others, or even count against. So, suppose that at dinner I have most reason to order the vegetable tartine whereas you have most reason to order the quiche; based on our general preferences and current desires, the fact that the tartine is spicy is a reason for me to order it and a reason for you to avoid it. Or that I have most reason to pursue a career in design while you rationally ought to pursue one in research. Or, arguably, that my particular moral weighting of tact over honesty gives me most reason to offer feedback in one way while your alternative ordering gives you most reason to phrase it differently.

In each of these cases, by hypothesis, the orchestra of 'world-given' reasons leaves the case indeterminate, but the contribution of 'identity-given' reasons tips the scale. On the dogmatic model, when faced with a choice we either do as we already have most reason to do (that is, we act in conformity with our extant identity, plus the balance of world-given reasons) or we do not act at all. The only way to move through time and space, when we are moving as we should, is in compliance with this identity. Surely we are not *born* with an identity complex enough to contain the operating instructions for a lifetime, and so a minor puzzle presents itself: if the dogmatic rationalists are right about how humans choose, it seems that we are likely to be trapped from infancy in rational paralysis, unable to behave in any way not dictated by a balance of reasons which requires the contribution of identity-based reasons, yet without the identity that could ground those reasons.[230]

Only the most rigid dogmatist can be caught by this puzzle, however. The classical model itself will let us transcend the *tabula rasa* of infancy—vaulting from choice to choice under the steam of animal instinct and a rational policy of exploring the unknown, learning new preferences from those experiences to guide future choices, quite possibly supplementing those modes of self-transformation with non-action-centric modes like reflection and brute causal force—without any break in the sort of rational guidance it makes essential to human agency. What it won't permit, I argue below, is the conscious control of such choices: *deliberate* self-transformation.

Some forms of self-transformation, to be sure, may be justified by the shape of one's existing identity. Suppose I am a smoker, but would prefer to be a non-smoker, because

[230] For more on this point, see Millgram 1999.

regular smoking makes my throat itch and my hands smell of tobacco even after washing. Simplistically, we may say that I have one nexus of first-order reasons that condones smoking (arising from nicotine addiction, an appreciation of the space for calm reflection or socialization that smoking enables) and another that counts against it (visceral aversion to certain aspects of the experience, knowledge of associated health risks). Perhaps this conflict generates a second-order reason to abandon smoking and that is actualized in choice; perhaps the battlefield is only that of the reasons already mentioned and the non-smokers have it; in either case, a transformation in one of the habits that is a part of my practical identity is seemingly made possible without any gap in rational guidance.

Or is it? Self-transformation of this kind occurs because of some existing, identifiable pressure in one's practical identity that presses against the status quo. It is not so much a transformation of the self as it is an actualization of an identity that has not yet been fully instantiated. If the above is correct, and the balance of reasons truly points away from smoking, I am already, with respect to the reasons that arise from my identity, a non-smoker. All that's left is acting as one when the opportunity presents itself.

Genuine self-transformation, it seems, must have occurred at some point prior to the one identified above. The self changed when its nexus of identity-based reasons shifted from one favoring smoking to one favoring the opposite. There must have been some moment where the balance of reasons favored smoking—even if only via the rational policy of trying new things just once—and therefore, it would seem, there must have been some liminal moment when the balance of identity-based reasons neither favored nor disfavored

smoking. That is to say, there must have been a moment of underdetermination lurking in the past.

We need presume no strict continuity in normative status here—that a straight line from 'smoke' to 'don't smoke' must cross the x-axis at least once, or anything of that kind—for there are a variety of ways that the underdetermination might have manifested. Perhaps my fear of cancer is suddenly augmented by the smoking-related illness of a beloved teacher, so that the balance quickly tips [i]. Perhaps my need for reflective breaks from work slowly diminishes as I find other opportunities for quiet meditation outdoors; I wake up one morning and realize that my desire to smoke is no longer quite the match for my non-smoking inclinations [ii]. Or perhaps for a while, with respect to smoking at least, I have really been not one but at least two selves—smoker and non-smoker—competing for actualization, but I *identify* with one over another and therefore wind up with most reason to act for the reasons it offers and devalue those of the other [iii].

The first story is the only one consonant with dogmatic versions of the classical theory, since it is the only one that features uninterrupted rational guidance. In [i], a sudden transformation of my desires and preferences tips the rational balance, flipping the switch on smoking as it goes. In such cases, self-transformation is basically hydraulic. It is difficult to believe that many episodes of self-transformation can be assimilated to this model. We need only consider situations like [ii], where the transformation is more gradual, or [iii], where the shift is characterized by internal conflict rather than an abrupt discontinuity between internally coherent selves (smoker one minute, non-smoker the next), to see

alternative models whose basic structure simply does not admit of seamless rational guidance.

A classical accounting of the rational structure of [ii] and [iii] would hold that the moments of incomparable or incommensurable choice between selves must be rationally-arbitrary blips rather than deliberate, comparatively self-intelligible movements of the will.

In [ii], a gradual transformation of my desires leads me to a point of indifference, at which point I must make a rationally arbitrary decision: will I smoke again today, or not, given that I have equally compelling reason to do either? If I flip a coin and opt to cease smoking, perhaps I'll enjoy the experience of not-smoking so much that I gain a new reason (or strengthen existing ones) to persist in this way of being. The transformation, again, will have occurred on the basis of existing identity-based reasons, aside from a brief moment of underdetermination, and while the path does not exhibit unbroken rational governance, the shift in identity occurs in a way that requires no agential input beyond a randomly-directed impulse—smoke or not today?—that might have gone either way.

In [iii], the choice between competing existing selves (or fragments thereof), whether we view it as exhibiting indifference or incommensurability, could only ever admit of a similar resolution. The exercise of will that transcends my deliberative impasse is comparatively unintelligible, more akin to the arbitrary acts of will described by the **sheer voluntarist** than any which route willed action through personal identity.

So the classical model of choice can accommodate *diachronic* shifts in personal identity, but not *deliberate* ones. And hydraulicism is once againt the source of the problem: the static relation between the acting self and its identity-based reasons combined with the

mechanistic relation between the balance of reasons and the acting self ensures that no account of deliberate self-transformative choice is possible. When we change, whether from within or without, we seem to be acting either as marionette or hand puppet rather than self-directed rational agent.[231]

On our map, the **normative voluntarists** represent the first effort to give an account of self-transforming choice which admits some diachronic fluidity into the identity–reasons nexus. They retain the fixed relation between identity and reasons—whatever the self is, its reasons are the source of comparative rational justification and therefore of intelligibility in choice—but allow that the relation between the acting self and the balance of reasons is bidirectional. When the balance of reasons is perfectly even, or even unreadable, the will can alter that rational balance by shifting the self conception that underwrites the identity-based reasons already on the scale. Yet trying to unpack the normative mechanics of this bidirectional process—our identity determines our reasons, but our will can alter our identity in order to change those reasons—led us quickly into confusion, and ultimately to the collapse of the model.

The **radical voluntarists** offer a more serious take on the diachronic nature of personal identity and rational choice, one whose mechanics are less obviously problematic—if only because they are more opaque. For the **heroic existentialist**, the self is no more nor less than the sum of its choices; our identities play no role in fixing the balance of reasons prior to choice (nor does the world outside us), but they are an important means of understanding and ultimately affirming the pattern of choices our lives exhibit. On the one

---

[231] Paraphrase of Nozick 1981, p. 310, *supra* n. 51, §**3**.

hand, this notion of personal identity takes seriously the idea that we can be profoundly shifted by our choices, and that the self that emerges from the choice is partly forged by the choice itself. (Indeed, for the old heroes, the self that emerges from choice is *entirely* forged by it; all choices are, by definition, self-launching.) For the **constructivist**, self-constituting choice requires deliberation on a sense of identity prior to and during choice-making; the intelligibility of choice resides in its coherence with a self-conception endorsed by the agent who inhabits it in the moment of decision.

The kind of bad faith that we identified in §4, however, rested in the temporal sleight-of-hand that each account involved. Self-launching or self-affirming choice, if it is to be made intelligible by the identity of the acting self, required that some of the relations between identity and the balance of reasons exhibit a perplexing retrospective reorganization, or synchronic 'p & $\neg p$' structure. These amounted to *mauvaise foi* because it seemed impossible that a self-conscious agent could accept contradictions of this kind without undermining the intelligibility and therefore the meaning of any such choices. An agent cannot see herself as both bound and not bound by her existing identities in the same moment; cannot see the meaning of her past choices as expressing an identity (set of commitments and judgments) which by definition did not exist at the moment the choice was made.

But what if we've gotten it wrong about personal identity because everyone so far surveyed has been operating with an artificially frozen reference frame? What if the various voluntarist claims described above are insights so central to the notion of self-conscious personal identity that we must reframe our ideas about rationality to suit them? If the

nature of the self is, as the radicals have it, "not being but becoming,"[232] perhaps we haven't gone far enough in constructing a theory of normative facts that pays due deference to the ineliminably diachronic nature of reflective selfhood. This, anyway, is how we invert the methodological assumption that agency rightly precedes identity in assessing views of rational choice: if we can integrate the diachronic nature of reflective selfhood *into* our model of intelligible agency, perhaps we can build a model of underdetermined choice-making which explains how the self interfaces with the choice in a way that renders such choices rationally intelligible even as they are authentic moments agent-authored self-transformation. What might this model look like?

Here is one, which builds the temporally-complex Heideggerian notion of the 'always–already' into the stories of self-launching choice and retrospective intelligibility embraced by the heroic existentialists and constructivists.[233] In a 1971 essay on ideology, political philosopher Louis Althusser develops a theory of normative instantiation—of the process by which norms are legitimated for individuals in social practice—which accords normative properties the right kind of temporal flexibility. On Althusser's view, ideological imposition (the process by which norms gain legitimate authority over subjects, or certain normative properties are instantiated) occurs via a process of *misrecognition*:

> [I]deology … imposes (without appearing to do so) obviousnesses as obviousnesses, which we cannot fail to recognize and before which we have the

---

[232] Paraphrase of Nietzsche, *The Will to Power*, §708 (*supra* n. 65, §**4**); *cf*. Korsgaard 2009, §1.1.5 (p. 7).

[233] Here is Hegel scholar Catherine Kellogg on the 'always–already': "…at every new moment in the movements of Mind, we find Spirit or Mind 'always–already' at its next moment. Thus, sensible intuition emerges as always–already attention, attention emerges as always–already imagination (and so on)" (Kellogg 2006, p. 213).

> inevitable and natural reaction of crying out (aloud or in the 'still, small voice of conscience'): 'That's obvious! That's right! That's true!'[234]

The insidiousness of ideology is that it presents itself as 'always–already' binding upon us.[235] We view ideological claims not as proposals for consideration but as obviously already so, and our misrecognition (*méconnaissance*) of their normative status unwittingly makes true their claim to bind us. Althusser calls this process of "ideological recruitment," by which we are transformed from free individuals to subjects of the ideology in question, *interpellation*, a process analogous to that by which we hail a friend in the street:

> *…all ideology hails or interpellates concrete individuals as concrete subjects, by the functioning of the category of the subject.* …[a] very precise operation … which can be imagined along the lines of the most commonplace everyday police (or other) hailing: 'Hey, you there!'[236]

Our response to this hail—*you there!*—constitutes recognition of the legitimacy of the claim addressed (*yes, officer, I am the one you seek*); our failure to respond to mis-directed hails (*hi, George!*) signals rejection of the claim that we are as the addresser claims we are (*I am not George*). For Althusser, these responses do not simply signal recognition of normative legitimacy (or illegitimacy), they *constitute* that legitimacy.

---

[234] Althusser 1971.

[235] Addressing his reader, Althusser notes that "you and I are *always already* subjects, and as such constantly practice the rituals of ideological recognition" (1971, emphasis in original).

[236] Althusser 1971, emphasis in original. Fuller selection: " ideology 'acts' or 'functions' in such a way that it 'recruits' subjects among the individuals (it recruits them all), or 'transforms' the individuals into subjects (it transforms them all) by that very precise operation which I have called *interpellation* or hailing, and which can be imagined along the lines of the most commonplace everyday police (or other) hailing: 'Hey, you there!'"

Interpellation is a conceptual rather than a temporal process, and so the misrecognition of ideological claims as binding cannot be undone simply by recognizing one's own mistake. When a friend hails me in the street and I return his greeting, the exchange occurs in sequence. But when it comes to ideological interpellation,

> in reality these things happen without any succession. The existence of ideology and the hailing or interpellation of individuals as subjects are one and the same thing.[237]

This is the sense in which recognition of ideological claims is constitutive of their authority rather than caused by it. Ideology is eternal, writes Althusser, because the 'obviousnesses' presented by ideology seem to have always–already been there.[238] It is essential to the nature of ideological claims that their normative status exhibits this temporal peculiarity; the nature of the process by which something which seems always–already to have been true of everyone becomes true for a particular individual at a particular time is strangely instantaneous.[239]

---

[237] Althusser 1971.

[238] "As ideology is eternal, I must now suppress the temporal form in which I have presented the functioning of ideology, and say: ideology has always-already interpellated individuals as subjects, which amounts to making it clear that individuals are always-already interpellated by ideology as subjects…" (Althusser 1971).

[239] I've skimmed over a number of additional elements to this picture, including the key claim that ideology is required for subjectivity in the first place, since they are not essential to the structural aspects of the view I'm interested in applying to the problem of underdetermined choice-making. For more on the view, see Dolar 1993, Purvis and Hunt 1993.

Interpellation thus exhibits a kind of mythic structure, argues Rebecca Kukla, and it is not the only process of normative legitimation with this form.[240] In a lengthy essay on the normative status of epistemological claims in Wilfrid Sellars' work, she identifies a myth-making pattern of discourse in which claims about the past serve to legitimate normative facts in the present without any pretense of factual accuracy. The myth of Prometheus, for instance, is not offered as literally true by those who invoke it to caution against scientific overreach; yet invoking it is meant to (at least partially) legitimate the normative claim that hubris is to be avoided. According to Kukla, claims about our past perceptions of "normatively inflected" facts (her examples are epistemological) sometimes serve "a regulative rather than reporting role,"[241] their aim not to represent how things were at the time but how they must seem to us now, given who we have become:

> …the act of remembering, like the act of recognizing, can function as a constitutive demand … [one's] memories are *achieved products* of this process rather than found contents of [one's] mind. This is possible because what memory demands … is not a change in or a creation of empirical facts, but rather a certain normative stance towards 'past' experiences and reports; [we] demand of [our] memories that they live up to the claims of the present.[242]

Like the Althusserian hail-and-response model, the act of recalling past events—say, choices—is not a simple act of perceiving what was normatively true at the time, but

---

[240] Kukla 2000. My debt in this section to Rebecca's ideas is obvious, but I want to explicitly call out both her personal warmth and terrifyingly brilliant 2000 essay, both of which I am glad to have in my life. Her essay, which I encountered in my very first year of graduate study, is about so much more than its title suggests; it holds a high spot in a very short list of philosophical works whose impact on me as a reader was so forceful as to feel, at times, like physical assault. These are the moments a philosopher lives for.

[241] Kukla 2000, p. 192; *cf.* pp. 187–188.

[242] Kukla 2000, p. 189.

demanding that whatever we use to make sense of the choice—say, a particular self-conception—was always–already there by misperceiving it as such. This, then, might be one way of recasting the existentialist idea that retrospective intelligibility is a legitimate ground of normative facts.

Recall the *mauvaise foi* of the radical voluntarist project: each variety examined seemed to require that certain claims ('*x* is the thing there is most reason to do') be both true and not true of a given choice. For the existentialist, a self whose identity only emerges over the course of a lifetime can be used to give choices this classical form—'*x* was the thing there was most reason to do'—because there being 'most reason to do' a thing is reduced to its being in some sense inevitable or inescapable for the self forged by it. For the constructivist, affirmation of an extant identity makes sense of the choice in the moment it is performed, but this affirmation takes place in the face of—indeed explicitly requires—the fact that this identity is not inescapable but in fact might be shucked off at any moment. In both cases, something about the choice made is meant to secure its own inevitability, therefore its own intelligibility, therefore its own rationality.

The existentialist picture was found to be in bad faith because its account of retrospective intelligibility was too inflexible to accommodate any notion of 'identity' rich enough to play the required role, and its model of rationality too derivative of willful self-launching to deserve the name. Kukla's myth-making model opens up a third possibility: that our claims about past normative facts need not match up tidily with our past perception of those facts in order to play a legitimating role in the stories we tell about ourselves. Could

if the *self*—and correlative notions of *self-conception* or *identity*—be, in this very specific sense, a myth?

The voluntarist effort to make sense of underdetermined choices—whether prospectively or retrospectively—was driven by a need to find meaning in the choices we make that are not dictated for us by the world that confronts the identity we already inhabit. The normative voluntarists and constructivists aimed at prospective recalibration of the balance of reasons, so that classical intelligibility was on hand for us prior to the moment of choice; the existentialists gave up on that entirely and threw their lot in with retrospection. Retrospective recalibration of the balance of reasons—granting us authority to alter the normative terrain after the fact—seemed as impossible, if not more so, than the in-the-moment recalibration hoped for by the normative voluntarist, and indeed we found the heroic existentialists endorsing a somewhat more complicated picture of how retrospective recalibration works.

This additional complication came in the form of a radically re-envisioned model of rational intelligibility: when the balance of reasons is sufficiently unfamiliar, any number of claims about how its verdict is shaped may seem equally (or incommensurably) plausible. Our new proposal would retain a classical understanding of what the balance of reasons *is*, and offer a rather more complicated model for how its verdict may be altered over time. In particular: given that identity-based reasons, as already assumed, can and often do tip the rational balance, if our identities are themselves essentially in flux, but in a structured and comprehensible way, there might be a level of understood underdetermination—sounds like **indeterminacy**, a shifty value in the rational calculus—that characterizes any choice

where one's self-conception is implicated. If claims about the self, or rather, claims about our the self's practical footprint, serve a "regulatory rather than reporting role" in the stories we tell about how and why our lives unfolded as they did, perhaps the radical voluntarist project is not doomed to double-speak.

Consider the role played by myth in Kukla-style stories of normative constitution:

> In order for us to be able to follow the mythical imperative, the myth must express a coherent possible history… if the myth asks us to mis-recognize our past for certain purposes of legitimization, then we do not want to either assert its literality (which would just be a mistake) nor to deny it (which would undercut its ability to serve its pragmatic purpose). Rather, we must maintain a stance … wherein we actively *keep open* the possibility that the myth is true, but without *asserting* its truth.[243]

If self-mythologizing is a tool for legitimating past choices, it must partially constitute the normative facts that justify choice in the first place: facts about the balance of reasons, including identity-based reasons.

5.2 Redeeming the assumption

Suppose self-mythologizing is possible. When does occur? I invent an identity for myself: say, the always–already philosopher. Is that myth of use to me in now making sense of the choice I made years ago to ditch ballet for graduate school, or must I have marshaled it to explain the choice *while* I was making it?

[243] Kukla 2000, p. 199.

Let's try retrospection first, echoing Kukla's suggestion that our memories of normatively-inflected facts must measure up to the demands of our current perspective.[244] Six years ago, I made a seemingly underdetermined choice to pursue a degree in philosophy. As the years wear on, the imagined paths of forgone opportunities become fainter, choked with weeds. If I come to tell a story in which the balance of identity-based reasons at that moment in fact favored the choice I made—in a way that is not strictly factually-accurate but *might* have been true, such that I can "keep open" the possibility of its veracity without committing to it—well, isn't this a familiar epistemic move? I certainly have enjoyed philosophy; surely one explanation for this state of affairs is that I was always meant to do it, more suited to it all along than ballet, or architecture, or medicine. The myth of myself as the 'always–already philosopher' provides a tidy explanation for the choice, one that frames my sense of self moving forward as well as looking back. Who's to contradict this version of events? Mightn't one's own identity be precisely the sort of thing that is partly constituted by our own perception of it, before, during, or after the fact?

On this model, myth legitimates choice by misrecognizing the self who made it. When I tell the story of that choice to myself or others, I do so in terms that privilege the path taken because the self I recall is so colored by the self I now am that this misperception actually recalibrates the normative balance.

Now, an essential part of the Althusserian model (and of Kukla's epistemic one) is that the hailed party—another person, our own perceptual memories—can meet the demand lodged by interpellative misrecognition. In each case, what is misrecognized is the *authority*

---

[244] Kukla 2000, p. 189.

of some normative claim or other over the individual in question, not the content or truth value of the claim itself. For self-mythologizing to turn underdetermined choices contexts to determined ones after the fact, to turn misrecognition from error to normative constitution, there must be a some *exchange* that effects the shift. But diachronic identity does not exhibit this form, for the simple reason that our past selves have no means of responding to our misrecognition of them, and so the authority of normative facts based on a misrecognition of the past can't utilize the power of myth in the same way.

For Althusser, the misrecognition of ideology's claims as 'always–already' regnant in our social world constitutes them as authoritative because the normative contours of that world don't have an existence separate from the subjectivity of its, well, subjects. Whatever else is true of the model, it has this mechanism built in. For Kukla, the misrecognition is creative and surprising, since what is misrecognized is not the authority of a cluster of social norms but the authority of certain claims about past perceptual experience, which seem much less amenable to constitutive misconstrual on first glance. But this is only because it is natural to conceive of epistemic facts like 'it looked green to me at the time' as facts whose authority in further epistemic reasoning depends essentially upon their truth-value. Kukla rejects this notion. In the cases of misrecognition she examines there is no determinate character that the past perception has independent of the role it plays in "structuring the space of reasons and the epistemic standing of the experiencer,"[245] my claim to have had a particular kind of experience before I was a capable negotiator of the normative space required to place that experience in the right sort of inferential relations to other epistemic claims does not so much *conflict* with the past as it does *make sense* of it

[245] Kukla 2000, p. 180; *cf.* pp. 187–188.

using tools I only now have to comprehend it.[246] The myth is a source of normative warrant only because the normative arena about which it tells a story is structured by the necessity of such myth-making. Ordinary epistemic claims do not get off the ground without it as a foundation, just as there is no ideology without the sort of subjectivity which could sanction it. Misrecognition is a hail completed by the context in which it is made.

Retrospective claims about our identity do not manifest the same kind of responsiveness: we commonly make claims about our prior selves which leave the past untouched, and it is hard to see what it could be about identity that places it in the kind of peculiar normative space required, as we have already seen. The felt inescapability of certain choices *ex post facto* can easily be explained in terms of cognitive bias (as we saw in §4.3) or even in terms of having lost some of the language I once used to understand the world and the choice that I made then (see §3.5), neither of which is sufficient on its own to alter the balance of reasons.[247] Without this means of completing the process of constitutive misrecognition, then, we have no explanation for the authority of myth, and the retrospective account collapses.

---

[246] For the record, I have no view on Althusser's model of ideology and subjectivity, and I'm quite persuaded by Kukla's account of the misrecognition of *certain kinds of memories*. Since I don't think the model works for our purposes irrespective of whether it works elsewhere, I haven't spent any time defending it, and have spent only as much time unpacking it as is useful for seeing the precise way in which it fails for our needs.

[247] New research on memory does, indeed, indicate that our memories are almost literally re-written when called back to consciousness, though this has no obvious normative implications for our area of concern. See: Schiller and Phelps 2011, Quirk et al. 2010.

Perhaps, then, myth-making occurs contemporaneously with choice. Here again, myth legitimates choice by misrecognizing the identity of the person who performed it, but now misrecognition occurs simultaneously with choice-making. The role of the self in making choice intelligible is more like that envisioned by the constructivists than the existentialists. Since the authority of our identity in determining the balance of identity-based reasons is precisely what's at stake for the constructivist proposal, the myth-making model seems better suited to the task. Moreover, it seems we can avoid the problem of responsiveness that felled the existentialist version of the proposal: if self-misrecognition occurs in the moment, our identity isn't obviously rendered inert by temporal distance.

It is, however, rendered impotent by the structure that undercut the constructivist proposal in §4.5: that we must see our identities as at once binding-because-instantiated and not-instantiated-because-radically-up-to-us. Noting that the authority of the mythologized self depends on our accepting it as 'always–already' one way rather than another adds nothing to the original proposal, which seems to require much the same thing already. If I determine, in the self-launching moment, that in my heart I was all along committed to the version of myself for whom graduate school is the rational choice, either I was right (and the choice was never really underdetermined) or I am only pretending this is so (and so haven't moved the normative dial a whit, and am in bad faith besides).

And so our effort to accommodate the temporal complexity of reflective self-hood has come to naught. Its failure is, I think, instructive, for it demonstrates why other attempts to explain meaningful underdetermined choice by complexifying the self who acts are also likely doomed. The question of how deliberate self-transformation is possible is, as we

have seen, not very different from the question of how meaningful underdetermined choice is possible. The failure of the self-mythologizing proposal took on a familiar cast because the fault lines are the same here as they were before: no matter how rich a starting ground your temporally-smeared conception of personal identity provides, you still owe an account of the intelligibility of choice which invokes a contemporaneous balance of reasons, and so the same questions must be answered. A new notion of personal identity cannot furnish us with the tools needed to close the gap between intelligibility and arbitrariness; all it can ever do is push discussion into a parallel zone of discourse marked by the same fissures and fault lines. No new habitable area on the map of conceptual possibilities has been opened.

5.3 Assumption two: conceptual

We move, then, to the second assumption hypothesized to ground the intractability of the classical–voluntarist debate: adherence to a comparativist standard of rational choice. Comparativism has been the one fixed point in the dialectic so far. There is good reason for this—as I shall argue, comparativism about rational choice is not just intuitively plausible but also theoretically powerful, exerting a kind of gravitational force on competing alternatives which causes them to collapse, before too long, into something that looks an awful lot like comparativism.

In this section, we examine the dominant alternative to comparativism as a rational standard: satisficing. Like comparativism, the satisficing model of practical rationality is based on compliance with the balance of reasons, but it offers a more expansive criterion

of what compliance involves. From the beginning, we have traced a dialectic that presumes that rational intelligibility co-travels with comparative assessment—but if reasons make choice intelligible in ways other than by contributing to a comparative balance of merits, then much else about the debate must be reconsidered.

It is generally agreed that satisficing, as a theory of rational choice, came into prominence in the middle of the twentieth century as a result of Herbert Simon's effort to ameliorate a peculiar schism in social scientific models of human decision-making. Whereas economists and political scientists are wont to attribute "a preposterously omniscient rationality" to the agents they study, with complete and coherent sets of ranked preferences and infinite computational resources, he argues, psychologists and sociologists typically "reduce all cognition to affect" and instead model the ways in which social pressures and biological impulses drive choice.[248] Surely, he reasoned, these models should inform one another. Human rationality is bounded, and our models of rational decision-making must *at least procedurally* reflect this reality, without abandoning the possibility of rational standards for choice altogether. And so Simon's theory identifies all that "real human beings, of bounded rationality… faced with complexity and uncertainty," can do: "lacking the wits to optimize [with respect to preference satisfaction], they must be content to *satisfice*."[249]

---

[248] Simon 1976, pp. xxvi–xxvii.

[249] Simon 1979, p. 3, emphasis mine.

There are various ways of formalizing the satisficing standard of rationality as an alternative to the comparative model.[250] Simon's own work is often ambiguous[251] between two aims—descriptive and normative—and further complicated by slides between different versions of the normative claim. As a descriptive psychological theory, satisficing is true but (for our purposes) normatively uninteresting: we commonly satisfice with respect to our options, and do so consciously and deliberately. But satisficing exists too as a normative psychological theory, which holds that we *ought* commonly to satisfice for the sake of our own mental health.[252] Insofar as it is rational to pursue our own mental health or well-being, there is of course a derivative sense in which a satisficing orientation toward the world is rational, a legitimacy that might be inherited by particular token instances of satisficing choice. Alternatively, satisficing might show up as a normative theory about the *morally* appropriate deliberative attitude: that a certain kind of humility, flexibility, or acceptance might be one we have compelling moral reason to adopt.[253] In this way, satisficing as a global perspective (and manifested in token instances) might turn out to be rational too—but this again is a claim about what external reasons (moral, prudential) we have to adopt an orientation rather than a claim about the basic structure of rational choice-making. These views are basically orthogonal to our concern, but can easily complicate discussions of satisficing. Our focus is not on whether satisficing manifests

---

[250] See Byron 2005 for a useful overview.

[251] Byron 1998, p. 71n has a nice discussion of this ambiguity. Most philosophers, e.g. Richardson 2004, Schmidtz 2004, assume the descriptive reading of Simon's work even if they address a normative conception of satisficing itself.

[252] Haidt 2006.

[253] Slote 1986, 2001; Swanton 1993.

some particular psychological or moral virtue, but whether it can bear weight as a legitimate alternative to the comparativist model of practical rationality.

As a normative claim in rational choice theory, then, the satisficing view counts us as rational when we select an option judged to be 'good enough' even when we do not judge it to be the best available. This formulation admits of a *de dicto* and a *de re* reading: we might be rationally permitted to choose suboptimally when we know *that there exists* some superior option; or we might be rationally permitted to choose suboptimally when we know that our selection is inferior to *some particular option*. Defenders of what I will call *pure satisficing* endorse the latter; *tempered satisficing* the former. Below, we examine each with an eye to the intelligibility-making relationship each describes between reasons and the options they favor. I show that each version unravels upon examination, leaving us with no viable alternative to the comparativist standard that does not also abandon some very basic ideas about what reasons are and how they operate in rational choice.

5.4 Redeeming the assumption

Proponents of satisficing tend to argue from cases: from our unwillingness to call a particular action irrational, even when that choice is made in full knowledge that the selection is suboptimal, we may infer that the standard of rationality to which we all implicitly appeal is a satisficing, rather than an optimizing one. Consider, for instance, this classic case:

moving house You are selling your home, and your aim in setting your list price is neither to maximize profit, nor optimally balance profit against time on the market, but simply to get what you judge to be a satisfactory, or 'good enough' price for it.

Michael Slote, a leading proponent of pure satisficing, argues that the protagonist of **moving house** is not irrational to set some satisfactory threshold, prior to action, and stop searching for buyers so long as she encounters one that will meet her price, even if she might, by staying on the market for a bit longer, turn an even higher profit on the sale.[254]

It is not, Slote hastens to add, that the agent in question is implicitly balancing the utility of expected profit against other considerations—anxiety about time on the market, stress about courting buyers, mental effort put into setting the price 'just so,' or anything of this nature—he is simply:

> …content with good enough and does not seek to maximize (optimize) his expectations. His desires, his needs, are moderate, and perhaps knowing this about himself, he may not be particularly interested in doing better for himself than he is likely to do by selling at a merely satisfactory price. If someone pointed out that it would be better for him to get more money, he would reply, not by disagreeing, but by pointing out that for him at least a good enough price is good enough.[255]

Slote traces our real disinclination to fault the house-mover to the fact that rational standards are, in fact, satisficing rather than robustly comparative, and gives another example:

afternoon snack You're sitting at work next to a cache of free snacks. You'd enjoy a candy bar and a drink, wouldn't ruin your appetite for dinner if you had these, and could easily snag one—but you don't.

---

[254] Slote and Pettit 1984, p. 142. The example appears often in Slote's work; he explains that he has adapted the example from discussions of satisficing in economics.

[255] Slote and Pettit 1984, p. 143.

Here again, Slote insists, there's nothing irrational about your action, even though you had good reason to spring for the snack, and no compelling reason to refrain from doing so. A comparativist description of your action—choosing an option (not eating the candy bar) which you *know* to be inferior to some alternative (eating it) which is already clearly in view—would deem it paradigmatically irrational behavior, which intuition tells us it is not.

But is that really so? Slote goes on:

> You may simply not feel the need for any such snack. You turn down a good thing, a sure satisfaction, because you are perfectly satisfied as you are. Most of us … are not boundless optimizers or maximizers, but are sometimes (more) modest in our desires and needs. But such modesty, such moderation, need not be irrational or unreasonable on our part.[256]

The apparent success of both examples rests, I think, on a simple confusion between *maximizing a certain quantity* (of money, of physical satisfaction) and *optimizing with respect to all the reasons one faces in acting.* Consider the candy bar ascetic in **afternoon snack**: if she truly is "perfectly satisfied" as she is, then she does *not* have any compelling reason to head to the fridge; if her desires are truly "modest" then it's hard to see in what sense she would actually enjoy the snack. Someone with simple tastes would, naturally, lack compelling reason to actively seek to maximize sensual pleasure, so it seems that Slote's example and his intuition-pumping justifications are mutually destructive.

So too with our home-seller in **moving house**. Recall Slote's appeal to self-knowledge: "his desires… are moderate, and perhaps knowing this about himself, he may not be particularly interested in doing better for himself than he is likely to do by selling at a

[256] Slote and Pettit 1984, p. 144.

merely satisfactory price."[257] Someone literally "not…interested in doing better for himself" might fail to maximize profit on the sale of the house, but is surely still *optimizing* with respect to the full range of considerations he weighs in choosing how to act, which naturally include his own disposition. Indeed, a tendency toward threshold-setting and 'good enough' attitudes in financial matters is a point of pride for many people, whether they view penny-pinching as inherently distasteful or worry that fixating on acquisitive details will distract them from nobler aims. Surely these considerations, whether moral or merely personal, properly carry weight in the deliberative economy of a rational agent. Having compelling moral or prudential reason to adopt a certain kind of attitude toward choice is not the same thing as operating within a rational framework with an essentially satisficing structure. Think of the rational calculus that renders the satisficing attitude morally or prudentially rational: is that, too, satisficing in nature? It must be for the strong claim about rational standards to be true, but nothing in the forgoing has given any argument for it.

In fact, I believe we can show that whatever intuitive plausibility Slote's examples (and others like them) have is actually parasitic on, rather than subversive of, a comparative standard of rational choice. In each case we understand the reasonableness of the agent's choice in terms of its optimality *with respect to a full range of considerations*, even though each represents a failure to *maximize* a quantity—money or pleasure—that most agents value. Let's begin by tweaking the first example so as to shift it into the category of *de re* or pure satisficing:

---

[257] Slote and Pettit 1984, p. 143.

moving house$_2$ You are selling your home, and receive two offers in one day: one from a buyer wants it at the listed price—a price you'd be pleased to receive—and another offering more than twice what you're asking.

If our protagonist chooses the lower offer over the higher because they're both 'good enough' (so there's no compelling reason to choose one over the other), it's exceedingly hard to see in what sense his action can still be called rational—*unless* it reflects some deep-seated self-conception based on thrift or humility. But if it truly reflects these ideals, then the choice might not be suboptimal after all: it's the optimal choice *for this agent*, because of the values the choice deliberately expresses, even though it fails to maximize an otherwise desirable quantity. So either the choice is truly suboptimal, in which case (it seems to me) intuition no longer bears out an assessment of the choice as 'rational,' *or* it is optimal after all, in which case the intuition is in line with a comparative standard of rational choice.

So, too, for the **afternoon snack**. Tweaking the example to include some of the justificatory material Slote includes, and to eliminate whatever work is being done by the decadence of Slote's proposed snack (a candy bar and Coca-Cola), we find a similar dilemma:

afternoon snack$_2$ You're sitting at work near a cache of free teabags. You'd enjoy a cup of tea and an invigorating stroll around the office, and wouldn't ruin your appetite or disrupt your work if you went for it—but you don't, because you feel perfectly content as you are.

If you feel *perfectly content* as you are, then it's hard to see how getting up to fix yourself some tea could possibly represent an improvement on your current condition. Yet it's precisely the fact that the snack *would* represent an improvement—one you choose not to make—that enables Slote's claim that you're failing to optimize with respect to the merits

of the case. It simply can't be both at the same time. And if we still hew to the intuition that staying at your desk is rationally permissible even though you'd enjoy and benefit from a short respite (having eliminated any noise introduced by an obvious disconnect between the immediate pleasure of sugary snacks and the longer-term displeasure they typically cause), it can only be because we see some *value* in a policy of not distracting yourself from your work by chasing after tiny improvements in material well-being. Yet whatever value this policy of moderation has—whether moral or prudential—would again be what's carrying normative weight in deliberation, as fodder for an optimizing calculus of merits. There is no need to appeal to a non-comparative standard of rational choice in these cases, which do not stand up to sustained examination.

For many, the cases claimed above for pure satisficing were never convincing in the first place. Far more popular is the less strict *de dicto* understanding of the satisficer's claim that we are rationally permitted to knowingly make a suboptimal choice: we may do so in cases where we know *that* our selection is not the best, but do not know *to which* particular options it is inferior. This commonly occurs in choice contexts that are *dynamic* rather than *static.*[258] In a static choice context, all relevant alternatives are known, as are the reasons that favor them, and (barring indifference and indeterminacy) the balance of reasons is ready to hand. In dynamic contexts, by contrast, not all options have been unearthed and articulated at the most useful level of specificity, not all weights assessed. In such cases, satisficers

[258] Schmidtz 1995, ch. 2.

argue, it can be rational to choose an option one takes to be 'good enough' even though further deliberation will likely bring superior options to light.[259]

Most proponents of what I'll call *tempered satisficing* endorse it, following Simon, as a "humanly rational strategy"[260] appropriate for agents who "lack the wits to optimize" and must be content to approximate (without actually aiming at) a strong comparative standard in decision-making.[261] We commonly encounter situations in which we doubt that our choice is truly optimal, but have neither concrete knowledge of a superior alternative nor resources for further deliberation. It is often reasonable to employ a 'stopping rule' in deliberation: having set some threshold or 'aspiration level' for the choice in question, you stop searching for and weighing alternatives once you've found one that satisfies *this* criterion, rather than continuing the search. Tempered satisficing might well be a useful substantive principle for practical deliberation, and even a socially useful criterion for assessing choice. But is it really a standard in its own right if it boils down to "optimizing on a budget"[262] for creatures of bounded rationality?

David Schmidtz argues for a form of tempered satisficing whose normative warrant mostly derives from its utility as a globally optimific decision strategy, yet insists that it is a genuine alternative to the comparative model of rational choice, and not merely a version of it. 'Subtle optimizing' goes on, in Schmidtz's view, when one selects a 'good enough'

---

[259] Sen 1997, p. 475, labels this pressure to decide before all options are unearthed *decisional inescapability*.

[260] Schmidtz 2004.

[261] Simon 1979, p. 3.

[262] The phrase is from Narveson 2004.

option because it is the best of those options already enumerated at a point when the costs of further deliberation (time, mental energy) outweigh the expected incremental benefit of finding a superior option.[263] Genuine satisficing, by contrast, has its standard of 'good enough' set *internally* by the deliberator herself, rather than *externally* by limits on time and energy that constrain any finite agent—and it can often be rational to satisfice in this way.

Let's note here that even the hardest of hard-line comparativists will reject any deliberative procedure based on straightforward optimization that is self-undermining: some of the most broadly important goals in life, like happiness or meaningful relationships with others are precisely the kind one cannot aim at directly, as critics of simple consequentialism remind us.[264] In such cases, we can best attain the broad goal by making concrete choices, and adopting local deliberative strategies, which aim at different things entirely. The ultimate normative justification for these choices and strategies—why we ought to do those things and have these policies rather than others—will still be that they are the best means to the global goal. Should it come to light that some local policy is ineffective in this regard, comparativists say, we would face normative pressure to abandon it.

And this is true for most theories of tempered satisficing as a rational standard: insofar as their power to deem certain satisficing choices rational derives from the fact that adopting a satisficing strategy in the face of trade-offs between time, energy, and incommensurable goals helps us optimize with respect to a broader life plan by enabling an equitable spread of deliberative resources across choices, they are not interesting alternatives to the

[263] Schmidtz 1995, pp. 28–29. See also the appendix to Schmidtz 2004.

[264] Williams and Smart 1973; Kapur 1991.

comparative model. You have reason to settle for 'good enough' in any particular case *only if* it enables you to divert your attention—deliberative or otherwise—to other projects. We may even rationally choose to forgo an opportunity to rationally choose, if pausing to make that choice would use resources better spent elsewhere, as when, having slept through the alarm, we grab for the nearest outfit in the morning rather than selecting the best one to wear; or if the expected benefit from deliberation is tiny, as with the difference between a satisfactory can of soup or the best available; or if some fact about ourselves militates against deliberation in a particular case, as when, expecting an ex to a dinner party, we recognize that trying to figure out an optimal seating plan will spike what a low level of anxiety to a one requiring at least two glasses of wine before the guests even arrive. In each of these cases, not merely satisficing but actually *picking at random* would be rationally justified.

Comparativists, notably Michael Byron[265] and Philip Pettit,[266] have argued in this vein that a satisficing standard of rationality is unstable: however useful it might be as a strategy, as a formal criterion of choice its normative force is entirely derivative. Satisficing in contexts that where it is a *suboptimal* deliberative strategy—whether because we respond poorly to the (local) reasons at hand, or to (global) reasons we have to keep some time and energy in the tank for other choices—is irrational, and cannot be redeemed on its own terms. But Schmidtz argues, in response, that comparativist redescriptions of satisficing choice are "just-so stories" with no explanatory power that impute to agents a series of

[265] Byron 1998.

[266] Slote and Pettit 1984.

calculations about global optimization they are unlikely ever to have performed.[267] Further, he insists, satisficing is pervasive enough—not just experientially but conceptually—that insisting it can always be reduced to optimization is sheer dogmatism:

> There are cases where we do not care enough about the gap between the satisfactory and the optimal to make it rational to search for the optimal. Searching for the optimal toothpaste can be a waste of time, but so can searching for the optimal moment to quit looking for toothpaste. One way or another, satisficing enters the picture.[268]

Further, there will be times when even the most sophisticated optimizing strategies will be inappropriate, if they require information that may not be worth acquiring. Rational choice involves considering only those things that seem worthy of consideration, having a stopping rule that limits how comprehensive a body of information we must have before turning our attention to other matters.[269]

The charge of explanatory impotence would be an important one, if made to stick. If the comparativist response to rational satisficing is just an exercise in clever redescription by appeal to a global perspective no agent actually references, then it bears the same relation to reality as egoistic redescriptions of altruistic behavior: it's an unfalsifiable and therefore content-less theory. And just like egoistic redescription, the comparativist project is open to a charge of phenomenological inaccuracy. Many agents employing a satisficing strategy are unlikely to give a sophisticated justification for their reasoning in terms of global optimality; indeed, for some contexts putting an explicitly optimizing strategy out of one's

---

[267] Schmidtz 2004, p. 33.

[268] Schmidtz 2004, p. 37.

[269] Schmidtz 2004, p. 37.

mind is the very thing that makes satisficing the thing to do. The anxious dinner party planner, or any number of cases involving personal relationships or moral choices, are key examples where a consciously comparative attitude is *precisely* not for the best.

But all three of Schmidtz's charges here—the explanatory impotence and phenomenological inaccuracy of comparativist redescriptions, and the conceptual promiscuity of satisficing—are insufficient to redeem tempered satisficing as a rational standard. From the first-person perspective, there may be all the difference in the world between 'subtle optimizing' and threshold-based satisficing, or other deliberative heuristics that rational humans employ in pursuit of the ends of daily life. Yet in order to claim that their rational warrant derives from a comparative standard of choice, rather than belonging to the strategy itself, we need not demonstrate that the strategy was knowingly adopted by the agent because of its optimality, only that an agent would be rationally criticizable for clinging to the strategy were it brought to her attention that its outcomes were consistently suboptimal. And if this is so, then the charges of explanatory impotence and phenomenological inaccuracy fall together, with a direct assist from the last. If satisficing with respect to our satisficing (with respect to our satisficing…) is an inevitability—and I believe Schmidtz is right on this point—then we have all the more reason to grant the messiness of real-world deliberation, to admit the degree to which our satisficing choices are buoyed by hope rather than certainty, and to examine how these facts are already at work in our intuitions about specific cases.

Let's begin by reintroducing the example of the hurried morning dressing—and suppose that you have just a bit more time in the morning:

rushed morning You sleep through your alarm and roll out of bed with just a few minutes to get dressed. You've got a lot lined up for the day: a bike ride in to work, a presentation in the morning, two classes to teach in the afternoon, and a date in the evening.

Balancing all these considerations to select the optimal outfit—breathability and stretch on account of the cycling, professionalism for your presentation and approachability for class, not to mention just a touch of allure for the date after work—would be time-consuming, and time is something you're especially short on this morning. You spend a moment or two ruling out options impossible for cycling or inappropriate for class, then grab an outfit 'good enough' for the day's activities from those which remain. You've just made a choice that satisficed with respect to the optimality of your outfit, and optimized with respect to the broader concern of organizing the rest of your day (by seizing a few precious minutes to brush your teeth and check your tires before heading out).

Is it unreasonable to impute to you some judgment that satisficing was an optimific strategy in this context before employing it? And should your choice—which likely counts as 'subtle optimizing' in Schmidtz's lexicon—count as irrational if you did not?

The answer to both these questions is no. It is frequently reasonable to impute to people judgments they have not consciously made, and attitudes they are not aware they possess, on the basis of their actions and plausible counterfactual responses. Recall Schmidtz's point that in rational choice satisficing *always* "enters the picture" one way or another, even at the level of enumerating options in the first place.[270] For creatures of bounded time and rationality, '*keeping a little* [time, energy] *in the tank*' to deal with future decisions is a supremely rational policy. The most natural explanation for *why* it is rational is that it gives

[270] Schmidtz 2004, p. 37.

us the best chance of optimizing with respect to the full range of considerations—known and unknown, present and future—we need to address. Further, depending on personal disposition, having domain-specific policies of '*avoiding a maximizing attitude*' or '*embracing spontaneity*' can also be robustly rational, for reasons already discussed. But the limited capacities that make these policies reasonable in the first place are precisely what rule out the rationality of ascending to a truly global perspective to deliberate about deliberating—except in very rare high-stakes cases.

And so the comparativist must be careful in appealing to a global perspective, lest she impute to ordinary agents a rarefied form of meta-deliberation for any rational choice with a satisficing structure.[271] But she's also got an excellent explanation for *why* most rational agents don't occupy this perspective in normal conditions: we can often tell, over time, when a deliberative strategy is serving us poorly, and in the absence of such evidence it's reasonable to guess that our attention is better diverted to local projects. The same is true of a variety of heuristics you might employ, whether it's "sleeping on" important multi-variable choices (because of the article you read about sleep processing in *Psychology Today*), taking your mother's advice when it conflicts with your own judgment (on family matters but not interior decorating), trusting your gut about choosing business partners (so long as you've got a good track record on this front), and so on. Each of these represents a failure to directly ascertain and weigh all the reasons you face in certain contexts. They can be

[271] Schmidtz 1995 has a careful discussion of the tenability of global perspectives in deliberative choice, where he argues first that a global perspective is possible for normal agents, and then that this possibility does not set up an infinite regress of deliberative perspectives (where we must deliberate about deliberating about our deliberation). See also Smith 1991, and *cf.* Richardson 2004 for discussion of global and local perspectives orthogonal to this one.

adopted deliberately, as with the first-mentioned; can evolve over time, as is likely the case with the second, or can be strategies you aren't even really aware of utilizing, as with the third.

Suppose you stuck with your gut on business decisions in the face of overwhelming evidence that this was a poor strategy—it did not help you respond to the reasons you had to make certain partnerships rather than others, and the cost to your professional life was far greater than the satisfaction of an intuitive decision-making model. Would your actions still count as rational? After a point, surely they would not. Or suppose, after a string of failed ventures, you decided to move to a more analytical approach. Such a shift seems reasonable—because it's responding to a normative pressure to respond in the best way possible to first-order reasons that bear on particular decisions and second-order reasons to spend deliberative resources wisely.[272] So long as these procedures, whatever their etiology, are appropriately counterfactually responsive to failure, it's not unreasonable to attribute them to an agent and deem her actions rational in virtue of an implicit faith in their tendency to optimize over time.

Satisficers are right that we often quite reasonably do something other than optimize with respect to the full range of first-order reasons. But since our tendency to judge these cases 'rational' rather than irrational collapses in the face of evidence that those policies aren't appropriately responsive to evidence that they're suboptimal, we find that the standard of choice we used to judge them 'rational' all along could well have been a comparative one. The rational permissibility of satisficing lies in its instrumental value, always sensitive to

[272] I'm grateful to Ralph Wedgwood for helpful discussion on these points.

and derivative of the merits of the case at hand. This is true whether these are first-order reasons for choosing a particular option, or second-order reasons for choosing in a particular way—or opting not to deliberate at all.

I've chosen to dwell on satisficing as the primary alternative to a comparativist model of choice-making not just because of its contemporary predominance, but also because it neatly demonstrates two problems with abandoning comparativism. The first is that it's tricky indeed to find any instances of a non-comparativist standard in action that aren't easily re-described or counterfactually unraveled in a way that reveals the logical dominance of an optimizing, comparativist orientation toward choice. I take it that the bulk of the work above has shown this.

The second is less obvious, but ultimately more important. Every voluntarist in this work sought to articulate a notion of *meaningful* underdetermined choice, generating a variety of creative models of how reasons might work differently or intelligibility take a different form in underdetermined contexts than one might have supposed. They embraced a comparativist standard for rational, intelligible choice not because no one had thought of satisficing as an alternative account of rational standards, but because satisficing on its own says nothing about that issue. If true, it asserts the rational acceptability of non-optimal selections but gives us no grip on what it is that the intelligibility of those choices might consist in. As noted at the outset in §1, reasons are base currency in the economy of meaning that makes our lives comprehensible, and (if we are lucky) valuable. Our responses to the balance of reasons anchor the stories we tell that give our lives texture and meaning. Alternatives to comparativism that don't offer a novel theory of the way reasons

function in this economy will not offer any useful account of underdetermined choice even if they do not collapse under their own weight.

5.5 Conclusion

We opened this exploration with the apparently uncontroversial thought that rational choice aims at what is favored by the balance of reasons. We use reasons in deliberation for the purpose of comparison, weighing considerations against one another for the purpose of determining what to do. This view of practical reasoning is so widely held as to be virtually a truism; indeed, it is hard to understand what else reasons are *for* except to guide choice in this way. But this view seems ill-equipped to handle the theoretically possible (and probably ubiquitous) phenomenon of rational underdetermination. This was the observation which set our investigation in motion.

That investigation covered a range of possible positions in logical space: we were able to thematize competing theories of rational agency on the basis of certain theoretical treatments of underdetermination. The theories explored—one classical, two voluntarist—were united by their adherence to a comparativist model of practical reason and divided by the degree to which they took practical reasoning of this form to be the basis of meaningful human agency. We began with the classical theory of agency and slowly unraveled its commitment to a hydraulic model of practical reasoning—that is, slowly introduced volition into our picture of what matters in agency—finally abandoning that commitment in favor of a model of human agency in which volition, not rationality, is foundational. At no point along this journey did we uncover a satisfactory theory of

rational agency; indeed, at no point did we complete our quest for 'the missing agent,' since acts of the will turn out to be a poor proxy for the kind of creative, spontaneous, self-directed and self-comprehensible choices we have identified as meaningful exercises of agency.

We reached the edge of the map without traversing any truly habitable regions. The cost of taking up residence in the fiefdom of classical rationalism was too high; the stakes of living in more radical terrain at the end of our journey too much to bear; the hybrid middle kingdom more a hole in the map than anything else, a place not just uncomfortable but probably impossible to inhabit. That is to say: one may, it seems, be a dogmatic rationalist, but one's theory of agency is radically impoverished; one may be a radical voluntarist, but one's theory of rationality in choice barely deserves the description.

Rather than face the choice between evils, we have in this last chapter tried to broaden the map, to unravel two layers of presuppositions that might have obscured habitable terrain. Neither saved us. §5.1 showed how building our theory of rational agency around a complex notion of personal identity does nothing to alter the basic math that ties intelligibility to rational comparison; §5.2 showed how inescapable that math is even when we work very hard to open a space between the two. Changing methodology only resurfaced the fault lines of the original debate; trying to slip from the comparativist grip proved impossible without a robust theory of the intelligibility of choices that don't meet a comparativist standard.

Nothing about our exploration so far has shown that a robust theory of the intelligibility of choices which fall within some space of 'rational latitude' or (as I have called it)

'normative silence' is impossible. But it has shown, I think, how high a bar must be cleared in departing from the map. New accounts not just of rational intelligibility but of underdetermination itself will be called for, new modes of deliberate agency and new theories of personal development. To propose a tenable theory of rational agency, these alternative models must take care to avoid many of the mistakes uncovered here: mislaying the agent, reducing her to a contentless impulse, or accounting for her identity in overly-reductive terms. This is a bold move, more radical than any so far surveyed. Adding new implements to the deliberative toolbox will involve developing unfamiliar models of intelligibility (and therefore ultimately of meaning) in choice.

Departure of this kind is not unprecedented, but it takes us far past the boundaries of the current project and into uncharted territory. A diverse array of contemporary philosophers —Elizabeth Anderson, John Broome, Jonathan Dancy, Stephen Darwall, Patricia Greenspan, Maggie Little, Joseph Raz, Henry Richardson, to name but a few—are each in their own way exploring the space of formally underdetermined choice,[273] but have so far not arrived at a theory of rational intelligibility that meets the standards outlined in these pages. Their proposals involve pluralizing the kinds of normative force in the practical domain, or the kinds of reasons, or the kinds of relationship a rational agent can have to those reasons or forces, with the result that the space of rational underdetermination or of normative silence is a far noisier place than might have been classically understood. Just what that noise involves, though, and how we take it up in reasoning such that the choices we make there have the kind of intelligibility we know they need, are questions that must

[273] Anderson 1997; Broome 1997, 2001; Dancy 2004a, 2004b; Darwall 2001; Greenspan 2005, 2007, 2009, 2010; Little 2013; Raz 1975, 1988, 1997; Richardson 1990, 1997.

be answered before we can understand whether these other models are genuine alternatives to the views explored here. Nothing less than a new theory of rational agency is required.

If the arguments we've worked through here are to be believed, it is there we must turn for an account of rational underdetermination that does more than reflect back to our our own starting commitments. Finding no comfortable place on the map, we cast our eyes to the unfamiliar territory beyond, bearing in mind the cartographer's warning—*hic sunt leones*.

❧

Bibliography

Adams, Robert M. (1976) "Motive Utilitarianism," *Journal of Philosophy*, 73, pp. 467–481.

Althusser, Louis. (1971) "Ideology and Ideological State Apparatuses," in *Lenin and Philosophy* (London: New Left Books); available online.

Anderson, Elizabeth. (1997) "Practical Reason and Incommensurable Goods," in Chang, Ruth (ed.), *Incommensurability, Incomparability, and Choice*, pp. 90–109.

Anscombe, G.E.M. (1958) "On Brute Facts," *Analysis*, 18 (3), pp. 69–72.

Anselm, *De libertate arbitrii*, in Davies, Brian and G.D. Evans (eds.), *Anselm of Canterbury the Major Works*. (Oxford: Oxford University Press, 1998), available online.

Aquinas, *Summa Theologica* (New Advent Translation), available online.

Aristotle, *De Caelo*, available online.

Averroes, *Tahafut al-tahafut (The incoherence of the incoherence)*, trans. S. Van den Bergh. (London: Luzac, 1954).

Badhwar, Neera Kapur. (1987) "Friends as Ends in Themselves," *Philosophy and Phenomenological Research*, 48 (1), pp. 1–23.

Baier, Annette. (1985) *Postures of the Mind: Essays on Mind and Morals.* (Minneapolis, MN: University of Minnesota Press).

Bishop, John. (1989) *Natural Agency: An Essay on the Causal Theory of Action.* (Cambridge: Cambridge University Press).

Bratman, Michael E. (2000) "Reflection, Planning, and Temporally Extended Agency," *The Philosophical Review*, 109 (1), pp. 35–61.

Bratman, Michael E. (2001) "Two Problems about Human Agency," *Proceedings of the Aristotelian Society*, 101 (1), pp. 309–326.

Bratman, Michael E. (2002) "Circularity, Hierarchy, and Double Reduction," in Buss, Sarah and Lee Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, pp. 65–85.

Bratman, Michael E. (2003) "Autonomy and Hierarchy," *Social Philosophy and Policy*, 20 (2), pp. 156–176.

Bratman, Michael E. (2007) *Structures of Agency.* (Oxford: Oxford University Press).

Brennan, Tad. (2005) *The Stoic Life: Emotions, Duties, and Fate.* (New York: Oxford University Press USA).

Brink, David O. (1988) "Sidgwick's Dualism of Practical Reason," *Australasian Journal of Philosophy*, 66 (3), pp. 291–307.

Broome, John. (1997) "Is Incommensurability Vagueness?," in Chang, Ruth (ed.), *Incommensurability, Incomparability, and Choice*, pp. 67–89.

Broome, John. (2001) "Are Intentions Reasons? And how should we cope with incommensurable values?," *Practical Rationality and Preference: Essays for David Gauthier*, pp. 98–120.

Brunero, John. (2010) "The Scope of Rational Requirements," *The Philosophical Quarterly*, 60 (238), pp. 28–49.

Buridan, Jean. (ms) *Expositio textus of the De caelo.*

Burr, David. (1976) "The Persecution of Peter Olivi," *Transactions of the American Philosophical Society*, 66 (5), pp. 1–98.

Burrell, David. (1999) "Al-Ghazâlî on Created Freedom," *American Catholic Philosophical Quarterly*, 73, pp. 135–38.

Buss, Sarah. (1994) "Autonomy Reconsidered," *Midwest Studies in Philosophy*, 19 (1), pp. 95–121.

Buss, Sarah. (1997) "Weakness of Will," *Pacific Philosophical Quarterly*, 78 (1), pp. 13–44.

Buss, Sarah. (1999) "What Practical Reasoning Must Be If We Act for Our Own Reasons," *Australasian Journal of Philosophy*, 77 (4), pp. 399–421.

Byron, Michael. (1998) "Satisficing and Optimality," *Ethics*, 109, pp. 67–93.

Byron, Michael. (2004) "Could Aristotle Satisfice?," in *Satisficing and Maximizing: Moral Theorists on Practical Reason*, pp. 190–212.

Byron, Michael. (2005) "Simon's Revenge: Or, Incommensurability and Satisficing," *Analysis*, 65 (4), pp. 311–315.

Camoin, Francois. (1992) *Like Love, but Not Exactly.* (Columbia, MO: University of Missouri Press).

Cassirer, Ernst. (1953) *The Platonic Renaissance in England,* trans. James P. Pettigrove. (Austin, TX: The University of Texas Press).

Chang, Ruth. (1997) "Introduction," *Incommensurability, Incomparability, and Practical Reason*, pp. 1–34.

Chang, Ruth. (2009) "Voluntarist Reasons and the Sources of Normativity," *Reasons for Action*, pp. 243–272.

Chang, Ruth. (2013) "Commitments, Reasons, and the Will," forthcoming in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics*.

Chisholm, Roderick. (1966) "Freedom and Action," in Lehrer, Keith (ed.), *Freedom and Determinism*, pp. 11–44.

Chisholm, Roderick. (1976a) *Person and Object: A Metaphysical Study.* (La Salle: Open Court Publishing Company).

Chisholm, Roderick. (1976b) "The Agent as Cause," in Brand, Miles, and Douglas Walton (eds.), *Action Theory*, pp. 199–211.

Clarke, Randolph. (1996) "Agent Causation and Event Causation in the Production of Free Action," *Philosophical Topics*, 24 (2), pp. 19–48.

Clarke, Randolph. (2000) "Libertarian Views: Critical Survey of Noncausal and Event-Causal Accounts of Free Agency," in Kane, Robert (ed.), *The Oxford Handbook of Free Will*, pp. 365–85.

Cohon, Rachel. (2000) "The Roots of Reasons," *The Philosophical Review*, 109 (1), pp. 63–85.

Cooper, David. (1999) *Existentialism.* (Oxford: Blackwell).

Cudworth, Ralph. (1680) *Treatise on Eternal and Immutable Morality, with Treatise on Freewill.* (Cambridge: Cambridge University Press, 1996).

Dancy, Jonathan. (2004a) "Enticing Reasons," in Wallace, R. Jay, *et al.* (eds.), *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, pp. 91–118.

Dancy, Jonathan. (2004b) *Ethics without Principles.* (Oxford: Clarendon Press).

Danto, Arthur C. (1965) *Nietzsche as a Philosopher.* (New York: Macmillan).

Darwall, Stephen. (1995) *The British Moralists and the Internal 'Ought': 1640–1740.* (Cambridge: Cambridge University Press).

Darwall, Stephen. (2001) "Because I Want It," *Social Philosophy and Policy*, 18 (2), pp. 129–153.

DeBrabander, Firmin. (2007) *Spinoza and the Stoics: Power, Politics, and the Passions.* (New York: Continuum).

Debreu, Gerard. (1959) *A Theory of Value.* (New York: Wiley).

Della Rocca, Michael. (2008) *Spinoza.* (New York: Routledge).

Dellarosa, D. and L.E. Bourne. (1984) "Decisions and memory: Differential retrievability of consistent and contradictory evidence," *Journal of Verbal Learning and Verbal Behavior*, 23, pp. 669–82.

Dent, N.J.H. (1984) *The Moral Psychology of the Virtues.* (Cambridge: Cambridge University Press).

Descartes, Rene. (1641) *Meditations on First Philosophy.* ed. John Cottingham, trans. John Cottingham. (Cambridge: Cambridge University Press, 1985).

Dilhe, Albrecht. (1982) *The Theory of Will in Classical Antiquity.* (Berkeley, CA: University of California Press).

Dolar, Mladen. (1993) "Beyond Interpellation," *Qui Parle,* 6 (2), pp. 75–96.

Donagan, Alan. (1987) *Choice: The Essential Element in Human Action.* (London: Routledge).

Dostoyevsky, Fyodor. (1864) *Notes from Underground.* (New York: B.N. Publishing, 2001).

Fakhry, Majid. (1957) "The Classical Islamic Arguments for the Existence of God," *The Muslim World*, 47 (2), pp. 133–45.

Ferrero, Luca. (2010) "Decisions, Diachronic Autonomy, and the Division of Deliberative Labor," *Philosophers' Imprint,* 10 (2), pp. 1–23.

Fitzpatrick, William. (2005) "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity," *Ethics*, 115 (4), pp. 651–691.

Frankena, William. (1974) "Sidgwick and the Dualism of Practical Reason," *The Monist*, 58, pp. 449–467.

Frankfurt, Harry. (1971) "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy*, 68 (1), pp. 5–20.

Frankfurt, Harry. (1988a) *The Importance of What We Care About.* (Cambridge: Cambridge University Press).

Frankfurt, Harry. (1988b) "Identification and Externality," *The Importance of What We Care About*, pp. 58–68.

Franklin, Benjamin. (1956) *Mr. Franklin: A Selection from His Personal Letters.* eds. Bell, Whitfield J., Jr. and Leonard W. Labaree. (New Haven, CT: Yale University Press).

Gallagher, David. (1991) "Thomas Aquinas on the Will as Rational Appetite," *Journal of the History of Philosophy*, 29, pp. 559–584.

Garrett, Don. (2003) "Teleology in Spinoza and Early Modern Rationalism," in Gennaro, Rocco J. and Charles Huenemann (eds.), *New Essays on the Rationalists*, pp. 310–35.

Gert, Joshua. (2003a) "Requiring and Justifying: Two Dimensions of Normative Strength," *Erkenntnis*, 59 (1), pp. 5–36.

Gert, Joshua. (2003b) "Internalism and Different Kinds of Reasons," *The Philosophical Forum*, 34 (1), pp. 53–72.

Gill, Michael B. (2004) "Rationalism, Sentimentalism, and Ralph Cudworth," *Hume Studies*, 30 (1), pp. 149–81.

Greenspan, Patricia. (2005) "Asymmetrical Practical Reasons," in Reicher, M.E. and J.C. Marek (eds.), *Experience and Analysis: Proceedings of the 27th International Wittgenstein Symposium*, pp. 387–394.

Greenspan, Patricia. (2007) "Practical Reasons and Moral 'Ought'," in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics*, vol. 2, pp. 172–194.

Greenspan, Patricia. (2009) "Resting Content: Sensible Satisficing?," *American Philosophical Quarterly*, 46 (4), pp. 305–317.

Greenspan, Patricia. (2010) "Making Room for Options: Moral Reasons, Imperfect Duties, and Choice," *Social Philosophy and Policy*, 27 (2), pp. 181–205.

Griffin, James. (1997) "Incommensurability: What's the Problem?," in Chang, Ruth (ed.), *Incommensurability, Incomparability, and Practical Reason*, pp. 35–51.

Haidt, Jonathan. (2006) *The Happiness Hypothesis.* (New York: Basic Books).

Hare, R.M. (1981) *Moral Thinking: Its Levels, Method, and Point.* (Oxford: Clarendon Press).

Herman, Barbara. (2002) "Bootstrapping," in Buss, Sarah and Lee Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, pp. 253-74.

Hinchman, Edward. (2003) "Trust and Diachronic Agency," *Nous*, 37 (1), pp. 25–51.

Hobbes, Thomas. (1651) *Leviathan*. (New York: Hackett, 1994).

Hume, David. (1740) *A Treatise of Human Nature*. (Oxford: Clarendon, 2007).

Hume, David. (1748) *An Enquiry Concerning Human Understanding*. (Cambridge: Cambridge University Press, 2007).

Hursthouse, Rosalind. (1991) "Arational Actions," *The Journal of Philosophy*, 88, pp. 57–68.

Hutton, Sarah. (2007) "The Cambridge Platonists," in *A Companion to Early Modern Philosophy*, pp. 308–20.

James, Susan. (1993) "Spinoza the Stoic," in Sorrell, Tom (ed.), *The Rise of Modern Philosophy*, pp. 289–316.

Jauernig, Anja. (2008) "The Modal Strength of Leibniz's Principle: The Identity of Indiscernibles," *Oxford Studies in Early Modern Philosophy, Volume 4*, pp. 191–226.

Jones, Karen. (2008) "How to Change the Past," in Mackenzie, Catriona and Kim Atkins (eds.), *Practical Identity and Narrative Agency*, pp. 269–288.

Kahneman, Daniel, A. Tversky, and P. Slovik. (1982) *Judgment under Uncertainty: Heuristics and Biases.* (Cambridge: Cambridge University Press).

Kaufmann, Walter (ed.). (1954) *The Portable Nietzsche.* (New York: Penguin Books).

Kaufmann, Walter. (1974) *Nietzsche: Philosopher, Psychologist, Antichrist.* (Princeton, NJ: Princeton University Press).

Kant, Immanuel. (1781/1999) *Critique of Practical Reason*, trans. Paul Guyer and Allen Wood. (Cambridge: Cambridge University Press).

Kaye, Sharon M. (2004) "Why the Liberty of Indifference is Worth Wanting: Buridan's Ass, Friendship, and John Olivi," *History of Philosophy Quarterly*, 21 (1), pp. 21–42.

Kellogg, Catherine. (2006) "The Three Hegels: Kojève, Hyppolite, and Derrida on Hegel's Philosophy of Language," in Surbe, Jere O'Neill (ed.), *Hegel and Language*, pp. 199–234.

Kierkegaard, Søren. (1843) *Either/Or: A Fragment of Life*, trans. Alastair Hannay. (New York: Penguin, 1992).

Kierkegaard, Søren. (1843) *Fear and Trembling*, Evans, C. Stephen, and Sylvia Walsh (eds.). (Cambridge: Cambridge University Press, 2006).

Kierkegaard, Søren. (1846) *Concluding Unscientific Postscript to Philosophical Fragments*, trans. Howard V. Hong and Edna H. Hong. (Princeton, NJ: Princeton University Press, 1992).

Kolnai, Aurel. (1961) "Deliberation is of Ends." *Proceedings of the Aristotelian Society, New Series*, 62, pp. 195–218.

Korsgaard, Christine M. (1996a) *The Sources of Normativity.* (Cambridge: Cambridge University Press).

Korsgaard, Christine M. (1996b) "The Normativity of Instrumental Reason," in *The Constitution of Agency*, pp. 27–68.

Korsgaard, Christine M. (2009) *Self-Constitution: Agency, Identity, and Integrity.* (Oxford: Oxford University Press).

Kraut, Robert. (1986) "Love *De Re*," *Midwest Studies In Philosophy*, 10 pp. 413–430.

Kulstad, Mark. (1977) "Two Arguments on *Petites* Perceptions," *Rice University Studies*, 64 (4), pp. 57–68.

Kukla, Rebecca. (2000) "Myth, Memory and Misrecognition in Sellars' 'Empiricism and the Philosophy of Mind,'" *Philosophical Studies*, 101 (2/3), *The Philosophy of Wilfrid Sellars,* pp. 161–211.

Kukla, Rebecca, and Mark Lance. (2009) *'Yo!' and 'Lo!': The Pragmatic Topography of the Space of Reasons*. (Cambridge, MA: Harvard University Press).

Leibniz, Gottfreid Wilhelm. (1686) *Discourse on Metaphysics,* trans. George R. Montgomery. (2nd ed., New York: Open Court Publishing Company, 1978).

Leibniz, Gottfreid Wilhelm. (1704) *New Essays Concerning Human Understanding,* trans. Alfred Gideon Langley. (2nd ed., Chicago, IL: Open Court Publishing Company, 1985).

Leibniz, Gottfreid Wilhelm. (1710) *Theodicy,* trans. G.M. Duncan. (New Haven, CT, 1890).

Leibniz, Gottfreid Wilhelm. (1890) *Philosophical Writings.* eds. Mary Morris and G.H.R. Parkinson, trans. Mary Morris and G.H.R. Parkinson. (London: Dent, 1973).

Leibniz, Gottfreid Wilhelm. (1989) *Philosophical Essays,* trans. Roger Ariew and Daniel Garber. (New York: Hackett).

Leibniz, Gottfreid Wilhelm, and Samuel Clarke. (1715) "Exchange of papers between Leibniz and Clarke," available online.

Levi, Isaac. (1980) *Hard Choices: Decision Making under Unresolved Conflict.* (Cambridge: Cambridge University Press).

Levi, Isaac. (1982) "Ignorance, Probability, and Rational Choice," *Synthese*, 53, pp. 387–417.

Lloyd, Genevieve. (2008) "Shaping a Life: Narrative, Time, and Necessity," in Mackenzie, Catriona and Kim Atkins (eds.), *Practical Identity and Narrative Agency*, pp. 255–268.

Marcel, Gabriel. (1956) *The Philosophy of Existentialism,* trans. Manya Harari. (New York: Carol Publishing Group).

Mates, Benson. (1986) *The Philosophy of Leibniz: Metaphysics and Language.* (New York: Oxford University Press USA).

Mather, Mara, Eldar Shafir, and Marcia K. Johnson. (2000) "Misremembrance of Options Past: Source Monitoring and Choice," *Psychological Science*, 11 (2), pp. 132–38.

McCord–Adams, Marilyn. (1987) "William Ockham: Naturalist or Voluntarist," in Whippel, John F. (ed.), *Studies in Medieval Philosophy*, pp. 203–235.

McDonough, Jeffrey. (2011) "The Heyday of Teleology and Early Modern Philosophy," *Midwest Studies In Philosophy*, 35, pp. 179–204.

McDowell, John. (1998) "Virtue and Reason," in *Mind, Value, and Reality*, pp. 50–73.

Melamed, Yitzhak L. (2010) "The Metaphysics of the Theological-Political Treatise," in Melamed, Yitzhak L. and Michael A.Rosenthal (eds.), *Spinoza's Theological-Political Treatise: A Critical Guide*, pp. 128–142.

Melamed, Yitzhak L. and Martin Lin. (2011) "Principle of Sufficient Reason," *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition).

Melamed, Yitzhak L. (2012) *Spinoza's Metaphysics of Substance and Thought.* (Oxford: Oxford University Press).

Millgram, Elijah. (1999) *Practical Induction.* (Cambridge, MA: Harvard University Press).

Moss, Jessica. (2010) "Aristotle's Non-Trivial, Non-Insane View that Everyone Always Desires Things under the Guise of the Good," in Tenenbaum, Sergio (ed.), *Desire, Practical Reason, and the Good*, pp. 65–81.

Murphy, Mark. (2001) *Natural Law and Practical Rationality.* (Cambridge: Cambridge University Press).

Murdoch, Iris. (1999) *Existentialists and Mystics*, ed. P. Conradi. (New York: Allen Lane/ Penguin).

Nagel, Thomas. (1996). "Universality and the Reflective Self," in Korsgaard, Christine, *The Sources of Normativity*, pp. 200–209.

Nehamas, Alexander. (1985) *Nietzsche: Life as Literature.* (Cambridge, MA: Harvard University Press).

Nietzsche, Friedrich. (1886) *Beyond Good and Evil*, trans. Rolf-Peter Horstmann. (Cambridge: Cambridge University Press, 2001).

Nietzsche, Friedrich. (1908) *Ecce Homo*, in *Nietzsche: The Anti-Christ, Ecce Homo, Twilight of the Idols*, trans. Ridley Aaron and Judith Norman. (Cambridge: Cambridge University Press, 2005).

Nietzsche, Friedrich. (1882) *The Gay Science*, ed. Bernard Williams, trans. Josefine Nauckhoff. (Cambridge: Cambridge University Press, 2001).

Nietzsche, Friedrich. (1883–1885) *Thus Spake Zarathustra*, trans. Walter Kaufmann. (New York: Random House, 2011).

Nietzsche, Friedrich. (1968) *The Will to Power*, ed. Walter Kaufmann. (New York: Vintage).

Nozick, Robert. (1981) *Philosophical Explanations.* (Cambridge, MA: Harvard University Press).

O'Connor, Timothy. (1995) "Agent Causation," in O'Connor, Timothy (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, pp. 173–200.

O'Connor, Timothy. (1996) "Why Agent Causation?," *Philosophical Topics*, 24 (2), pp. 143–158.

O'Connor, Timothy. (2000a) *Persons and Causes: The Metaphysics of Free Will.* (New York: Oxford University Press, USA).

O'Connor, Timothy. (2000b) "Causality, Mind, and Free Will," *Philosophical Perspectives*, 14, pp. 105–117.

Olivi, Peter. (ms/1926) *Quaestiones in secundum librum Sententiarum*, ed. B. Jansen, available online.

Paphitis, Sharli Anne. (2009) "The Doctrine of Eternal Recurrence and its significance with respect to On the Genealogy of Morals," *South African Journal of Philosophy*, 28 (2), pp. 189–198.

Parfit, Derek. (1984) *Reasons and Persons.* (Oxford: Clarendon Press).

Parfit, Derek. (1988) "What We Together Do" (unpublished manuscript).

Patrides, C.A. (ed.). (1969) *The Cambridge Platonists.* (Cambridge: Cambridge University Press).

Pettit, Philip. (1991) "Consequentialism," in Singer, Peter (ed.), *A Companion to Ethics*, pp. 230–237.

Plato, *Meno*, trans. R. Allen, in *The Dialogues of Plato, Vol. 1*. (New Haven, CT: Yale University Press, 1989).

Plato, *Gorgias*, trans. R. Allen, in *The Dialogues of Plato, Vol. 1*. (New Haven, CT: Yale University Press, 1989).

Plato, *Symposium*, trans. R. Allen, in *The Dialogues of Plato, Vol. 2*. (New Haven, CT: Yale University Press, 1989).

Plato, *Protagoras*, trans. C.C.W. Taylor. (New York: Oxford University Press USA, 2009).

Pollock, Jackson L. (1983) "How Do You Maximize Expectation Value?," *Nous*, 17 (3), pp. 409–421.

Purvis, Trevor, and Alan Hunt. (1993) "Discourse, Ideology, Discourse, Ideology, Discourse, Ideology…" *The British Journal of Sociology*, 44 (3), pp. 473–499.

Quirk et al. (2010) "Erasing Fear Memories with Extinction Training," *The Journal of Neuroscience*, 30 (45), pp. 14993–14997.

Railton, Peter. (1988) "Alienation, Consequentialism, and the Demands of Morality," in Scheffler, Samuel (ed.), *Consequentialism and Its Critics*, pp. 93–133.

Raphael, D.D. (ed.). (1991) *British Moralists 1650–1800,* vol. 1; (New York: Hackett).

Raz, Joseph. (1975) *Practical Reason and Norms.* (Princeton, NJ: Princeton University Press).

Raz, Joseph. (1988) *The Morality of Freedom.* (New York: Oxford University Press, USA).

Raz, Joseph. (1997) "Incommensurability and Agency," in Chang, Ruth (ed.), *Incommensurability, Incomparability, and Practical Reason*, pp. 110–128.

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 (2), pp. 142–175.

Richardson, Henry S. (1990) "Specifying Norms as a Way to Resolve Concrete Ethical Problems." *Philosophy and Public Affairs,* 19, pp. 279–310.

Richardson, Henry S. (1997) *Practical Reasoning about Final Ends.* (Cambridge: Cambridge University Press).

Richardson, Henry S. (2004) "Satisficing: Not Good Enough," in Byron, Michael (ed.), *Satisficing and Maximizing: Moral Theorists on Practical Reason*, pp. 106–130.

Ross, M. (1989) "Relation of implicit theories to the construction of personal histories," *Psychological Review*, 96, pp. 341–57.

Rovane, Carol. (1998) *The Bounds of Agency.* (Princeton, NJ: Princeton University Press).

Rowe, William. (2006) "Free Will, Moral Responsibility, and the Problem of 'Oomph'," *Journal of Ethics*, 10, pp. 295–313.

Russell, Bertrand. (1900) *A Critical Exposition of the Philosophy of Leibniz.* (Cambridge: Cambridge University Press), available online.

Sartre, Jean-Paul. (1946) "Existentialism is a Humanism," available online.

Sartre, Jean-Paul. (1943) *Being and Nothingness*, trans. Hazel E. Barnes. (New York: Washington Square Press, 1993).

Scheffler, Samuel. (1985) "Agent-Centered Restrictions, Rationality, and the Virtues," *Mind*, 94, pp. 409–419.

Schiller, Daniela, and Elizabeth A. Phelps. (2011) "Does Reconsolidation Occur in Humans?" *Frontiers in Behavioral Neuroscience,* 5 (24), available online.

Schmidtz, David. (1995) *Rational Choice and Moral Agency*. (Princeton, NJ: Princeton University Press).

Schmidtz, David. (2004) "Satisficing as a Humanly Rational Strategy," in Byron, Michael (ed.), *Satisficing and Maximizing: Moral Theorists on Practical Reason*, pp. 30–58.

Schroeder, Mark. (2009) *Slaves of the Passions.* (New York: Oxford University Press USA).

Scotus, John Duns. (1686) *Opera omnia*, eds. B. Hechich, B. Huculak, J. Percan, and S. Ruiz de Loizaga, available online.

Scotus, John Duns. (1308) *Ordinatio*, ed. Thomas Williams, available online.

Scotus, John Duns. (1552) *Quaestiones subtilissimae super libros Metaphysicorum Aristotelis*, eds. Etzkorn, Girrard J., and Allan B. Wolter, available online.

Sebo, Jeff. (ms) "Constructivism about the Self."

Sen, Amartya. (1992) *Inequality Reexamined.* (Oxford: Clarendon Press).

Sen, Amartya. (1996) "Justice and Assertive Incompleteness," delivered as the second Rosenthal Lecture at Northwestern University Law School, 1998.

Sen, Amartya. (1997) "Maximization and the Act of Choice," *Econometrica*, 65 (4), pp. 745–779.

Sen, Amartya. (2002) "Maximization and the Act of Choice," *Rationality and Freedom*, pp. 158–205.

Seneca, *Epistulae Morales*, available online [Latin].

Setiya, Kieran. (2007) *Reasons without Rationalism*. (Princeton, NJ: Princeton University Press).

Setiya, Kieran. (2010) "Sympathy for the Devil," in Tenenbaum, Sergio (ed.), *Desire, Practical Reason, and the Good*, pp. 82–110.

Shakespeare, William. (1609) *Sonnets*, available online.

Sidgwick, Henry. (1874) *The Methods of Ethics.* (7th ed., Indianapolis, IN: Hackett, 1981).

Simon, Herbert. (1976) *Administrative Behavior*. (3rd ed., New York: The Free Press).

Simon, Herbert. (1979) *Models of Thoughts*. (New Haven, CT: Yale University Press).

Simplicius. *Commentaria in Aristotelem Graecia, vol. 7, Simplicii in Aristotelis de Caelo Commentaria*, ed. I.L. Heilberg. (Berlin: Royal Prussian Academy, 1894), available online.

Slote, Michael and Philip Pettit. (1984) "Satisficing Consequentialism," *Proceedings of the Aristotelean Society*, 58, pp. 165–176.

Slote, Michael. (1986) "Moderation, Rationality, and Virtue," *The Tanner Lectures on Human Values*, 7, pp. 53–99.

Slote, Michael. (1989) *Beyond Optimizing: A Study of Rational Choice.* (Cambridge, MA: Harvard University Press).

Slote, Michael. (2001) "Moderation and Satisficing," in Millgram, Elijah (ed.), *Varieties of Practical Reasoning*, pp. 221–236.

Smart, J.C.C. (1973) "An Outline of a System of Utilitarian Ethics," *Utilitarianism For and Against*, pp. 3–74.

Smith, Holly. (1991) "Deciding How to Decide: Is There a Regress Problem?," in Bacharach, M. and S. Hurley (eds.), *Foundations of Decision Theory*, pp. 194–219.

Smith, Matthew Noah. (ms) "Bootstrapping and the Authority of Intentions."

Smith, Michael. (1994) *The Moral Problem.* (Oxford: Wiley–Blackwell).

Spinoza, Baruch. (1677) *Ethics*, trans. R.H.M. Elwes (1883), available online.

Spinoza, Baruch. (1991) "Treatise on the Emendation of the Intellect," in *Treatise on the Emendation of the Intellect and Selected Letters,* trans. Samuel Shirley. (2nd ed., New York: Hackett).

Spinoza, Baruch. (1670) *Theological–Political Treatise*, ed. Jonathan Israel and Michael Silverthorne. (Cambridge: Cambridge University Press, 2007).

Stonestreet, Erica K. (2008) "Self-Creating Reasons: The Normative Implications of Identity," Ph.D. dissertation, The University of Michigan Press.

Street, Sharon. (2008) "Constructivism about Reasons," in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics*, pp. 207–245.

Street, Sharon. (2010) "What is Constructivism in Ethics and Metaethics?," *Philosophy Compass*, 5, pp. 363–384.

Street, Sharon. (ms) "Coming to Terms with Contingency: Humean Constructivism about Practical Reason," forthcoming in Lenman, James and Yonatan Shemmer (eds.), *Constructivism in Practical Philosophy*.

Strong, Tracy B. (1975) *Friedrich Nietzsche and the Politics of Transfiguration.* (Berkeley, CA: University of California Press).

Swanton, Christine. (1993) "Satisficing and Virtue," *The Journal of Philosophy*, 90, pp. 33–48.

Taylor, Richard. (1966) *Action and Purpose.* (Engelwood Cliffs, NJ: Prentiss–Hall).

Taylor, S.E. and J.D. Brown. (1999) "Illusion and well-being: A social psychological perspective on mental health," *Psychological Bulletin*, 103, pp. 193–210.

Tversky, Amos, and Daniel Kahneman. (1991) "Loss aversion in riskless choice: A reference-dependent model," *Quarterly Journal of Economics*, 107, pp. 1039–1061.

Vailati, Ezio. (1990) "Leibniz on Locke on Weakness of Will," *Journal of the History of Philosophy*, 28 (2), pp. 213–228.

Velleman, J. David. (1989) "Epistemic Freedom," *Pacific Philosophical Quarterly*, 70, pp. 73–97.

Velleman, J. David. (1992) "What happens when someone acts?," *Mind*, 101 (403), pp. 461–481.

Velleman, J. David. (2006) "The Self as Narrator," *Self to Self*, pp. 203–223.

Velleman, J. David. (2007) "What good is a will?," in Leist, Anton (ed.), *Action in Context*, pp. 193–215.

Wallace, R. Jay. (1999) "Three Conceptions of Rational Agency," *Ethical Theory and Moral Practice*, 2 (3), pp. 217–242.

Wallace, R. Jay. (2006) *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason.* (New York: Oxford University Press, USA).

Watson, Gary. (2002) "Volitional Necessities," in Buss, Sarah and Lee Overton (eds.), *The Contours of Agency: Essays on Themes from Harry Frankfurt*, pp. 129–159.

Watson, Gary. (2004) "Asserting and Promising," *Philosophical Studies*, 117, pp. 57–77.

Williams, Bernard and J.C.C. Smart. (1973) *Utilitarianism: For and Against.* (Cambridge: Cambridge University Press).

Williams, Bernard. (1981) "Internal and External Reasons," *Moral Luck*, pp. 101–113.

Williams, Thomas. (1995) "How Scotus Separates Morality from Happiness," *American Catholic Philosophical Quarterly*, 69, pp. 425–445.

Williams, Thomas. (2011) "Human Freedom and Agency," in Stump, Eleonore and Brian Davies (eds.), *The Oxford Handbook of Aquinas*, pp. 199–208.

Wolter, Allan Bernard. (1997) *Duns Scotus on the Will and Morality.* (Washington, DC: Catholic University Press).