Syracuse University

SURFACE

Dissertations - ALL

SURFACE

May 2015

Where the Reasons Come From

Aaron Wolf
Syracuse University

Follow this and additional works at: https://surface.syr.edu/etd

Part of the Arts and Humanities Commons

Recommended Citation

ABSTRACT

We have a problem understanding our normative reasons. If there's a reason to φ, we can explain why that reason counts in favor of doing it with the fact that φ-ing would be good in some way. But it's also explainable by non-evaluative properties that make φ-ing good. Suppose the traffic gives us a reason to leave early for the airport. The deeper reason might be the non-evaluative fact that the traffic would cause us to miss our flight, or it might be the value of making the flight. Although both reasons are perfectly good at the everyday level, it would be wrong to say they both contribute to the fundamental normative story of why traffic favors leaving early. Because both are ways of calling attention to the same idea, that making the flight is our goal, they shouldn't be independent. One of these considerations is a reason only because the other is. If we considered each as its own self-contained reason we'd be double-counting. So which is the more fundamental one? I argue that values—goodness and badness of various kinds—give our best, most ultimate reasons.

WHERE THE REASONS COME FROM

By

Aaron Wolf

B.A. Muhlenberg College, 2005

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Philosohpy

Syracuse University
May 2015

Copyright © 2015 Aaron Wolf

ACKNOWLEDGMENTS

In the summer of 2007, the Syracuse philosophy department hosted a conference on practical reasons. I didn't go to that conference, but I did flippantly tell a group of friends who were going that I couldn't care less about reasons. With the benefit of hindsight, I think it's safe to say I was wrong about that.

Thanks to my committee members, especially to Daniel Star for being an external reader. Many thanks as well to the folks who read early versions of these chapters and gave helpful comments, including John Monteleone, Deke Gould, Hille Paakkunainen, and David Sobel. Special acknowledgement is due to the members of the "Executive Dissertation Council," who read nearly every word in here: Kirsten Egerstrom, Amy Massoud, Matt Koehler, Travis Timmerman, and Sean Clancy. Having the opportunity to talk through ideas with them was invaluable. The ninja-turtle theme is for them.

A very large portion of gratitude goes to Ben Bradley for helping me hang on to the good ideas, and for letting me empty my brain at him with little to no warning.

Most of all, I want to thank Ashley Taylor for her inexhaustible support, along with my parents, Kris and Pen Wolf, for everything. This dissertation is for them.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Reasons and wisdom

There's an old idea in philosophy that the practically wise do things for the right reasons. When I do something for the wrong reason—say, feeding my neighbor's pet turtles while he's away so that they live long enough to participate in the radiation testing I'm planning for them—I've exercised less wisdom than I could have, even though feeding them was the right thing to do. That much is commonplace. But philosophers tend not to notice the parallel conclusion that wisdom also requires understanding what those reasons are and where they come from. This is a dissertation about that overlooked virtue of human rationality: knowing the normative source of our reasons.

To see why wisdom requires this, consider Holmes and Watson solving a crime. Suppose they both conclude that the victim was poisoned, and do so on the basis of just one reason, namely that there is a broken wine glass near the body. But where Holmes understands that the glass provides a reason because it must have slipped out of the victim's hand as she was drinking, Watson has

developed an unreflective superstition about broken wine glasses and poisoning over the course of his work as a consulting detective and has no explanation of his inference. Surely Holmes has shown more intellectual virtue than Watson, and that's despite the fact that they have both responded to the exactly the same evidence. Or think of the religious believer who allows that the potential for harming others is a moral reason not to text while driving, and thinks that it's a reason because God frowns on causing harm to others. When she discovers the Euthyphro problem and realizes that God's approval can't be the right explanation of our moral reasons, she gains in practical wisdom what she has lost in suspect moral explanations. It should be clear from these examples that this kind of practical wisdom differs from the classical Aristotelian kind. Aristotle did not seem to require the wise to have a deeper understanding of where their reasons come from.[1] All the same, this other sort of wisdom is worth having too.

This dissertation defends a straightforward but controversial view about that kind of understanding: all of our reasons come from values. That is, our best, most ultimate reasons are considerations of value. Suppose for example that I face a choice between Job A and Job B, and that one reason for choosing B over A is location. The view says that the final explanation for why location considerations count in favor of B cannot rest with non-evaluative facts, like the fact that A is in the Northeast and B is on the West Coast. It must lie instead with something evaluative, like the fact that it would be better for me to be on the West Coast, or the fact that the cost of living compared against salary is no better at Job A than Job B. In other words, when there's a reason to do something, it has its normative force in virtue of some deeper reason

---

[1]See, e.g. Nicomachean Ethics, book 6.

featuring values, good or bad. Knowing that makes us more like Holmes and less like Watson.

1.2   The basic problem

If knowing where the reasons come from were easy, it wouldn't be philosophically interesting. There's a simple difficulty in accounting for the reasons we have. At first glance, both values and the qualities that make things valuable appear to succeed as reasons of the same kind. So, if traffic generates a reason to leave for the airport soon, that reason can be glossed pretty well with the fact that we need to be there in an hour and the traffic at this time of day usually makes it at least a 40 minute drive. But we can also give the reason by noting that it would be bad for us if the traffic caused us to miss the flight. We generally accept both kinds of considerations, value and non-value, at least outside of philosophical contexts. These explanations seem fine from the point of view of common sense, but trouble emerges when we realize that every reason appealing to non-evaluative facts competes with one involving evaluative facts, and vice versa.

For example, the fact that the ball landed next to the pin makes my golf shot a good one, but if I'm listing out the reasons I have for being happy with it, there'd be something odd about mentioning both its goodness and its proximity to the hole. Both considerations are ways of praising the result of the shot, so neither adds anything over the other. And the fact that I need to write a dissertation to get a job makes it good for me to write this from a professional standpoint, but its being necessary for a job and its being good for my career don't contribute to my reasons separately. These considerations

say effectively the same thing, so we risk double-counting if we allow both as independent reasons.

When this kind of competition arises, we're forced to make one of the reasons dependent on the other. Otherwise those of us who believe in normative reasons are playing right into the hands of skeptics and nihilists. We're talking about a kind of power after all, namely the ability that features of the world have to make it so that we should do and believe things. That's what normative reasons are about. Letting both values and makers give independent reasons in cases like this would be redundant and would multiply powers unnecessarily, which violates well-founded norms of theoretical simplicity. If that were the settled picture of reasons philosophy had to offer, they would look pretty suspicious.

So, we have a real question: are value facts the deeper reasons that underwrite the everyday reasons given by non-value facts, or is it the other way around? I don't think there is any antecedently obvious answer to this question. And because we go through life treating both as though they gave perfectly good reasons on their own, common sense looks unable to offer any clear guidance. Initial difficulties aside, though, I think the philosophical evidence ultimately supports values as giving the best and most ultimate reasons. That's what it means to say that reasons come from values. I'll try to show that in the chapters that follow.

1.3   Summary of what's ahead

The thought that reasons come from values is unpopular these days. The dominant view is that reasons come only from value-making properties,

made popular by T. M. Scanlon's buck-passing view of value. Folks who have recently defended versions of this idea include Derek Parfit, Jonathan Dancy, Mark Schroeder, and Jonas Olson, with historical antecedents in Franz Brentano, Henry Sidgwick, and A.C. Ewing. I'll give two arguments to the effect that this popular way of looking at things is wrong. But first we need some background.

Chapter 2 lays the dialectical foundation. We'll go into a bit more detail about the phenomenon of normative reasons and our basic problem accounting for them. The chapter also surveys the answers we can give to the problem, tracing out some of their history. I'll argue in particular that the view that reasons come from values has roots not only in the eudaimonistic outlooks of Plato and Aristotle, but in G.E. Moore as well. We'll also have a look at where the contemporary buck-passing tradition goes wrong, and why the famous Wrong Kind of Reasons problem for that view shouldn't trouble us here.

Chapter 3 delivers the positive argument for the idea that reasons come from values. The argument gets off the ground by combining two plausible and widely accepted principles: the maxim that *ought* implies *can* and the view that people are motivated only under the guise of the good. When you have a reason to do something, you rationally ought to do it, at least prima facie. If this obligation implies that the reason is something you can act on, and it's also true that we can be motivated only to the extent that we see some kind of value in what we're doing, then it follows that every normative reason ultimately comes from value. If there were a reason that wasn't in the end dependent on value, it would be unable to motivate and so it would violate Ought Implies Can. There are, of course, well-known objections to

both of these premises, but I'll argue that they can be weakened enough to avoid them while still implying that reasons come from values.

I'll suggest that we should accept some version or other of Ought Implies Can, because without it ethics loses an essential feature: the ability to guide our choices. If it weren't always true that you can, in some sense, do the things you should, then ethics stops being a discipline for real people's lives. At least one version of the principle is strong enough to work in the argument while dodging a number of the major objections. Roughly, this version says that nothing can be a reason unless it's psychologically possible for the person to be motivated by it. There's also good evidence for the Guise of the Good thesis. Most of the things we do we see some good in, even if we don't think very highly of them in the end. The primary complaints about the view tend to focus on goodness *per se* as a motivator, and they can be handled by modifying it so that we're not only motivated by the good, but by other values as well. These changes to the premises weaken the argument so that it can avoid the common objections while still showing that values are the best and most ultimate reasons there are.

Chapter 4 is about understanding the standard criticism of values giving reasons, which is best represented by Scanlon's redundancy argument. Because the qualities that make things valuable are sufficient on their own to produce the reasons we have—so the idea goes—values themselves are redundant because there's nothing left for them to do. All it takes is realization and a presumption that apparent competition should be resolved in favor of the lower level, and the higher-level realized things look useless. I'll argue that Scanlon has to be understood as making a particular kind of difference-making claim, namely that if we subtracted all the value from the world while holding

everything else fixed that we can, our reasons would remain unchanged. Without this sort of premise, the conclusion that values don't give reasons doesn't seem to follow. A survey of the alternative ideas bears out the need for this kind of difference-making premise. We'll call this the *Principle of Rational Exclusion*, which is rather attractive despite being misguided.

Chapter 5 goes on the offensive against the redundancy argument and criticisms like it, supplying the second argument against the idea that values don't give reasons. First we'll build on the understanding of redundancy as rational exclusion from Chapter 4 and address a few difficulties that the Principle of Rational Exclusion (PORE) might create. Then we'll turn to the argument itself: if PORE, then values don't give reasons, PORE; therefore values don't give reasons. Properly understood, both of its premises are doubtful. On the one hand, you might think that the Principle of Rational Exclusion shows us nothing about real-life reasons, since by making a prediction about what would happen if descriptive-to-normative supervenience were suspended, it's given up on a fundamental feature of normative reasoning and is therefore irrelevant. This counts against the if-then premise. On the other hand, if we grant that PORE could show something about our actual reasons, we risk it showing too much if we allow that it's true.

I'll show that the rationale behind PORE is structurally identical to the causal exclusion argument that gripped the mental causation debates in philosophy of mind during the 1980's and 90's. Once we see the similarities between the two inferences, we open up the possibility of modifying responses to the old causal exclusion argument into ways of resisting the new rational exclusion argument. In particular, Ned Block's rejoinder to causal exclusion shows promise. Block argued that if we agree that the mental is excluded

from having causal powers in favor of its non-mental realizers, then we must also agree that higher-level realizers must be excluded in favor of the next level down, and so on until the original causal powers we wanted to attribute to the mental drain away to the lowest level of physics, if there even is one. If there isn't a lowest-level, then causal powers drain away into nothing. I'll make an analogous point about values and their makers: if we presume that the makers are sufficient on their own to generate our reasons and exclude values on that basis, then we must also exclude higher-level value-makers in favor of lower-level ones, and so on until we reach the lowest level. Our reasons would then be given by properties that are too basic and complex for us to appreciate. If we can't appreciate the properties that really give our reasons then those reasons cannot be action-guiding, and therefore are not really normative reasons at all. So if we go in for rational exclusion, the reason-giving powers we ordinarily grant to macro-level properties will have drained away into nothing.

Chapter 6 addresses an important objection from chapter 3 in significant detail. Part of the positive argument for the view that reasons come from values is the Ought Implies Can principle. Perhaps the most common worry about Ought Implies Can is that it's incompatible with the widely-accepted thought that you can't get an *ought* from an *is*, known as Hume's Guillotine. The problem is that if obligation implies ability, then inability implies a lack of obligation. Since the claim that something can't be done looks like a straightforward *is* and the claim that it's not the case that it ought to be done looks like a straightforward *ought*, the two doctrines appear incompatible. I'll argue that despite its attractiveness, Hume's Guillotine cannot withstand scrutiny, and—fittingly—should be replaced with the Moorean view that there

is no synonymy between normative and descriptive terms, which is no threat to Ought Implies Can.

CHAPTER 2

THE RATIONAL COMPETITION

PROBLEM

As I suggested in chapter 1, we have a problem explaining where our reasons come from. On the one hand, the fact that something would be good to do appears to furnish a complete and independent explanation of why I have a reason to do it. And on the other, the very things that make it good to do also appear to give a complete and independent explanation. So, since I enjoy skiing, I have a reason to visit Utah for spring break. The fact that Utah has good skiing constitutes one perfectly good normative explanation for my reason to go there, one that doesn't seem to depend on anything else. But so does the low moisture content and high volume of the snow, which is what makes Utah skiing so great.

For realists about reasons, this is a problem because allowing both to give independent explanations is to needlessly multiply normative powers. Call this the *rational competition problem*. If we're going to take reasons seriously, they need to be well-behaved. That is, a realist theory of normative reasons

should be subject to the same constraints as other realist theories, which means avoiding things like widespread overdetermination or double-counting.

This chapter sets the stage for the action to come. The aim of the dissertation is to defend the idea that reasons come from values as a solution to the rational competition problem. But there's a lot contained within that slogan. What are normative reasons and why should we be realists about them? What kind of explanation of will a solution to the rational competition problem give us? What are the possible solutions? This chapter answers those questions, and addresses some skeptical concerns about the structure of normative reasons.

2.1 What are reasons?

Normative reasons are considerations that count in favor of doing something: taking an action or having a particular attitude. Here are some examples.

- That kale has health benefits counts in favor of eating it

- That kale tastes like dirt counts in favor of avoiding it

- That it would be good for me to finish my dissertation soon gives counts in favor of writing right now

- That it's my job to grade my students' papers in a reasonable time counts in favor of having them done by the end of spring break

- That positivity about the job market will make my life go better is a reason for remaining optimistic

- That the schedule says the bus will be by my house at 9:27 counts in favor of believing it

- That I can't cross the chasm and escape my pursuers without intending to counts in favor of resolving to do it

Normative reasons offer an explanation of the things they favor by justifying them. In other words, reasons explain in practical or rational terms why those things would be worth doing. These contrast with motivating reasons, which explain what led a person to do something. Here are a few cases of those.

- I bought a new bicycle because I felt my old one was beyond repair

- I texted Idelle because I wanted to set a time to meet for dinner

- I listened to U2 because it was St. Patrick's Day

- I believe there's a rat in the attic because of the noises I heard

When we act well, our motivating reasons coincide with our normative ones. But often they don't. We can get a nice illustration of how they differ by considering examples where I don't act well—that is, where my motivation conflicts with my normative reasons.

- I know that prudence requires me not to watch old episodes of "Teenage Mutant Ninja Turtles" on the internet until late at night, but I enjoy it too much to stop.

- I've promised to show up for a dissertation meeting at 9:30 this morning, but I didn't get much sleep last night and forgot, so I go out for breakfast unaware that I should be on campus instead.

- I've recently learned that football teams who intentionally down the ball by taking a knee win 80 percent of the time, and teams that do it twice

win 95 percent of the time. I confuse correlation with causation and conclude from this that I ought to direct the youth football team I coach to take a knee in each of its first two possessions.

These show us that normative reasons are the good ones, and the motivating are just the ones we use—good or bad. Motivating reasons will appear here and there, but the normative ones are the main focus of the dissertation.

The idea that normative reasons are considerations that favor actions or attitudes is called "the standard conception." For the purposes of this dissertation we'll be assuming that something like the standard conception is correct, though some might be skeptical. Those with misgivings about it should pay special attention to the last section of this chapter, where I try to address a range of concerns about the standard conception. For simplicity's sake I'll refer to the actions and attitudes considerations count in favor of as *objects* of reasons.

There is one thing worth saying about the standard conception here at the beginning. Philosophers are surprisingly loose in the ways they talk about reasons. The standard conception tells us that considerations give reasons, and considerations are typically understood as facts or true propositions. But when we're adhering to the vocabulary of the standard conception, we quite often talk about properties (or their instances) as giving reasons. I think we should be happy to stick with this looseness, but it needs to be made clear what these different locutions mean. Because normative reasons are explanations of why we should do or think something, the things that give the reasons need to have some kind of semantic content. Strictly speaking something like the standard conception is therefore unavoidable—true propositions seem to be what really count in favor of objects. Talk of properties giving reasons must

therefore be shorthand for something having to do with propositions, and I think we can capture it as follows. Properties (or property instances) give reasons just insofar as names for those properties appear in true propositions that count in favor.

This way of seeing properties as reason-giving has a noteworthy consequence. Suppose F is a name for property p, and G is also a name for p, and suppose further that $<\varphi$-ing would be $F>$ is a reason to φ. This means that p gives a reason to φ, insofar as it's F. Because $F = G$, we might be inclined to say that $<\varphi$-ing would be $G>$ is also a reason to φ, but that doesn't follow. To be a reason is to be an explanation, and not all co-referring terms have the same explanatory power, so P might give a reason in virtue of being F but not G. This is all just a roundabout way of saying that that semantic content matters in the business of giving reasons, and we can't substitute co-referring terms in reasons any way we like. This point will become important later.

I said above that reasons need to be well-behaved if we're going to be realists about them. But we haven't yet seen why we should want to be realists at all. The reason is simple: any evidence someone might give against believing in them would count as epistemic reasons, which are a kind normative reasons. So we're stuck with them, or at least anyone who cares about having support for their claims is. You might not believe in all the kinds of reasons mentioned in this dissertation—like moral ones—but just about everyone believes in some, including error theorists like J.L. Mackie, who thought that the claims of morality and evaluation were all false.[1] Even he had to give reasons for thinking that.

---

[1]See Mackie 1977.

Now, this last move might seem a bit quick. Sharon Street calls herself an anti-realist, not only about moral normativity, but the epistemic kind as well.[2] So she may seem like an example of someone who believes in no reasons whatsoever. It's important to bear in mind, though, that there's more than one way to be an anti-realist. One is to deny that there are any mind-independent truths about reasons, which is Street's view. But a much deeper kind is to deny that there are any truths about reasons at all—even mind-dependent or deflationary ones—and it's only this full-on nihilistic view that's being denied. The picture of reasons we'll be working with in the rest of the dissertation is compatible with very weak versions of truth-realism, like Street's constructivism. We needn't be platonists to accept it, nor do we need to be robust realists like Russ Shafer-Landau or David Enoch.[3]

2.2   Normative explanation for reasons

Solving the rational competition problem will tell us something important about reasons. But what? On the one hand, we might get an answer to the metaphysical question of what they are, or what they consist in. The standard conception of reasons as facts counting in favor of objects makes a first pass at this kind of explanation. Or we might get an answer to the question of how we have epistemic access to our reasons, or to normative facts in general. On the other hand, we can ask why it is that loose, everyday reasons have a genuine claim on us—that is, what are the most fundamental reasons we can give for doing something? Answering this question gives us what I'll call a *normative explanation of reasons.* That's the focus of the dissertation. In the

---

[2]Street 2006 and 2011.
[3]Shafer-Landau 2003 and Enoch 2011.

same way that the metaphysician might want to know what the end of the story for material composition is (simples, say), we are looking for the end of the rational story of why some object of a reason is favored.

Normative explanation of reasons appears to be of the same general kind as what people want when they ask "why be moral?" or "why be rational?" To see that it differs from the metaphysical or epistemic kinds, notice that giving an answer about what reasons consist in or how we can know about them won't satisfy. Someone with these questions wants to know why they should think morality or rationality have any force at all, or what the most gripping reasons are. This is an old point, but it's worth making again. Normative explanation for reasons works much the same way. Solving the rational competition problem will tell us which reasons are the most ultimate and most gripping reasons are, but that says very little about the metaphysical or epistemic questions.

Although our present concern is of the same kind as the "why be rational?" question, the goal is not to convince the dedicated skeptic that there is real normative stuff out there—some people just refuse to be gripped. But for those who accept at least some kind of normativity (and just about everyone should, as we've seen), the question of what its most basic form is can still arise: given that reasons compete, which are the more fundamental? That's our focus.

Observe that as with the skeptic's question, metaphysical and epistemic answers won't work. That reasons are considerations favoring objects says nothing about what our best reasons are. And the fact that we can know our reasons by knowing their realizers wouldn't tell us either—open question considerations should make both of these things clear. The normative

explanation of reasons is its own concern, and it should have a normative answer.

Why is this particular question worth resolving? Certainly not to win over folks who disapprove of all this reasons-talk. Rather, it's because we gain a kind of practical wisdom in knowing where reasons come from. As we saw in Chapter 1, rationality and intellectual virtue hold in higher esteem those who truly understand at a deeper level why they should do what their reasons demand. So to the extent that we want to know our world and what it tells us about the choices we should make, the normative explanation of reasons has a place.

This claim about the value of knowing where the reasons come from is parallel to a line that Daniel Star has advocated. Even if non-philosophers are able to identify and act on good reasons without simply getting lucky, there is a deeper truth to be had in finding the ultimate reasons for doing what we should. And this philosophical explanation is not a waste of time, it's wisdom worth having.[4]

2.3 Solutions to the problem

For every instance of value there are value-makers: properties in virtue of which things are good (and bad). Importantly, values and their makers need not be identical. Here are some examples.

- Values: being morally right, being a justified belief, being a funny joke, being imprudent

- Value-makers: being conducive to happiness, being a belief caused in the

---

[4]See Star 2011.

appropriate way, having a certain timing and delivery, being contrary to my long-term interests

As we've seen, values give straightforward, everyday reasons, and so do value-makers. At the everyday level, the low moisture content of Utah snow really is a reason to ski there, and it seems to provide just as complete an explanation of why we should ski there as the fact that it's good snow. Its being good doesn't add anything to the story. On the converse, once we explain that it's good, its good-making qualities don't add anything. Either would would be a complete snow-related reason on its own, and admitting both as independent reasons would lead to the familiar double-counting and overdetermination problems. The same goes for epistemic reasons. When the sudden dimming of the sky counts in favor of believing it will rain, it could be because dark clouds are good evidence for expecting rain or because they're strongly correlated with rain, and the same competition arises.

To be clear, the rational competition problem is not about which considerations are reasons at the everyday level. Both kinds are. But only one will provide the independent reasons, that is, the ones in virtue of which the other, related considerations get to be reasons. The problem is about choosing which group to privilege in the order of explanation.

Short of denying there's a problem in the first place, there are three possible responses. We could resolve the competition in favor of value-makers, in favor of values, or deny that there is a single, unified explanation of our reasons. I'll discuss all three in this section. The aim will be to give some background for each and to consider some objections.

Although each of the answers to the problem has some basis in well-known philosophical views, none of those views had anything like the rational

competition problem as their explicit target. They were meant to address other questions, which I'll try to distinguish in the discussion below. So it's not that people like Aristotle, Moore, or even Scanlon had exactly the same concerns in mind, it's that their ideas on similar issues surrounding normativity and reasons present a framework for understanding the ways of answering the problem before us.

2.3.1 Solution 1: value-makers

Consider first the idea that value-making properties, like the moisture content of Utah snow, give the best and most fundamental reasons. This view is a close cousin of what's known as Scanlon's *buck-passing* account of values and reasons. Here's Scanlon:

> What, then, are the relations between these natural [value-making] properties, the property of being valuable, and the reasons that we have for behaving in certain ways in regard to things that are valuable? There seem to be two possibilities. The first is that when something has the right natural properties it has the further property of being valuable, and that property gives us reason to behave or react in certain ways with regard to it. Moore seems to be taking this view about goodness when he says that it is a simple, unanalyzable, non-natural property. The alternative, which I believe to be correct, is to hold that being good, or valuable, is not a property that itself provides a reason to respond to a thing in certain ways. Rather, to be good or valuable is to have other properties that constitute such reasons ... It differs from the first alternative simply in holding that it is not goodness or value itself that provides reasons but rather other properties that do so. For this reason

I call it a buck-passing account.[5]

There are a few independent claims here, but focus for now on the one that says the properties of being good and being valuable don't give reasons, while value-making properties do. Although Scanlon doesn't engage in talk of normative explanation or what the ultimate reasons are, he clearly rejects the view that reasons come from goodness and other values, since he thinks they don't give reasons at all.

To get from the buck-passing view to the idea the that reasons come from value-making properties, we need to impose some additional detail (which Scanlon need not accept, but bear with me). For present purposes we are supposing that there are *levels* of reasons, and we need to do this because of the rational competition problem. If we don't, we risk double-counting our reasons. Once non-independent reasons have been divided into levels, the lowest level will be the one where the best and most ultimate reasons reside. Given that framework, Scanlon's account would come out as the view that value-making properties give these ultimate reasons, since they're the only properties that give reasons in the first place. And his argument against goodness giving reasons suggests that he's alive to the rational competition problem:

> First, when I consider particular cases it seems that these reasons
> are provided by the natural properties that make a thing good
> or valuable. So, for example, the fact that a resort is pleasant is
> a reason to visit it or to recommend it to a friend, and the fact
> that a discovery casts light on the causes of cancer is a reason to
> applaud it and to support further research of that kind. These

---

[5]Scanlon 1998, p. 97

natural properties provide a complete explanation of the reasons we have for reacting in these ways to things that are good or valuable. It is not clear what further work could be done by special reason-providing properties of goodness and value, and even less clear how these properties could provide reasons.[6]

It is of course open to Scanlon or any other buck-passer to deny that this argument appeals to competition (they might, for instance, think that goodness doesn't even *appear* to give reasons), or to accept the rational competition problem but deny that invoking levels of reasons is necessary. That's why the view that value-making properties supply the ultimate reasons is just a close cousin of buck-passing. But there's some reason to think many buck-passers would endorse this view, and Scanlon is certainly the inspiration for it. For the most part, I'll talk as if this just is Scanlon's view, keeping in mind the above caveats.

Now, the buck-passing account has two components. There's a *negative thesis* that says goodness never gives reasons while good-making properties do, and a *positive* one that analyzes goodness in terms of reasons—to wit, that goodness is the higher-order property of there being other non-value properties that do give reasons. To generate the idea that value-makers give the ultimate reasons from the previous paragraph, we only needed the negative thesis.

The negative thesis tells us that reasons are given exclusively by value making properties. On the view I'll defend here, values give reasons at both the ordinary and ultimate level. But in light of the examples of values we saw above, this dispute could look merely verbal. I'm inclined to call the property of being funny a value, but it's open to a buck-passer to call that a

---

[6]Scanlon 1998, p. 97

good-making property rather than a value itself. In that case, the only thing we would disagree about is whether to call it a value, making it a merely verbal dispute. We'd agree on all the normative questions as far as being funny goes.

Indeed, it might look like this is the case given Scanlon's claim that the values of rightness and wrongness not only give reasons, but very powerful reasons. Here's Scanlon again:

> The fact that an action would be wrong constitutes sufficient reason not to do it (almost?) no matter what other considerations there might be in its favor. If there are circumstances in which an agent could have sufficient reason to do something that he or she knew to be wrong, these are at best very rare. But if right and wrong always or even almost always take precedence over other values, this is something that requires explanation. How can it make sense, if we recognize values other than right and wrong and take them seriously, to claim that reasons of this one kind have *priority* over all the rest? I will refer to this as the problem of the priority of right and wrong over other values.[7]

This makes it seem like Scanlon thinks only that goodness *per se* fails to give reasons, while other considerations featuring different names for values do. This suggestion is confirmed by Scanlon's later claim that "more specific evaluative properties" often do give reasons.[8] If that's what Scanlon's view amounts to, there would be something to disagree about, but it wouldn't be much: one view says the fact φ-ing would be morally good (etc.) can be a consideration that favors doing it, and the other doesn't. That would be about it.

---

[7]1998, p. 148.

[8]Scanlon 2002, p. 513.

Yet in spite of appearances and Scanlon's way of putting things, I don't think this is the best way to understand the buck-passing account. For one thing, it's not clear that it will survive his own argument against goodness giving reasons, which we saw above. If the properties that make things valuable are sufficient on their own to generate reasons, then given that rightness and wrongness are values, whatever properties make things right and wrong should also be sufficient on their own. Thus there would be no work left for rightness and wrongness to do. Jonathan Dancy runs this argument the other way: if wrongness does give reasons over and above the ones given by wrong-making properties, then it seems like badness should too.[9] Either way we look at it, we are led to the conclusion that buck-passers shouldn't allow values of any kind to give reasons. And if those arguments weren't enough, S. Matthew Liao adds that if we think of the view as only barring "thin" evaluative properties like goodness from giving reasons, we lose one of the chief attractions of buck-passing, namely its ability to give a non-circular reduction of the evaluative realm.[10] Goodness would simply be reduced to other values on this version of Scanlon.

The appearance of a verbal dispute is likely due to the fact that values often have multiple names, and sometimes those names don't include the words 'good' or 'bad'. Thus we call jokes "funny" instead of "comedically good," and beliefs "justified" rather than "epistemically good." But we needn't conclude that being funny and being comedically good aren't the same thing. They are. So the buck-passer shouldn't call the property of being funny a value-maker. It seems to be a genuine value, because it's the metric by which we evaluate jokes. That can go a long way towards preventing the appearance of a verbal

---

[9]Dancy 2000.

[10]Liao 2010, pp. 427-431.

dispute.

David Chalmers has a test for verbal disputes. The test goes like this: we forbid any use of the term that's suspected of being the culprit and see if we can state the disagreement without it.[11] Now suppose the buck-passer continues to say that the property of being funny gives reasons. One natural way to explain the disagreement in that case would be to say that whereas my view takes that property to give reasons *in virtue of being a value*, the buck-passer would think it favors in virtue of something else, like a value-making property. But if we want to apply the Chalmers elimination test, that can't be it. If we're barred from using the word 'value', can we still find something to disagree over?

I think we can. Scanlon's primary rationale for the buck-passing view is the *redundancy argument* from above: since value-making properties would be sufficient on their own to give all the reasons we have, we should say that they really do give all the reasons, and value properties don't. Take the value-talk out of this and you get the schema that when the X would have been sufficient on its own to generate all the reasons we attribute to Y then the X give those reasons and the Y don't. That's a claim that people who think reasons come from values ought to be wary of, since it would require taking a stand on what things would be like at worlds where there were values without any value-making properties. Such worlds might not have any people or rational beings in them at all, so they may not have reasons either. What's more, I'll argue in chapter 5 that buying into this schema threatens to create a regress in which reasons drain away into unappreciable levels of ontology. So it appears there is something to disagree about that doesn't turn on uses of value-terminology. And it should be evident that it's at the center of the

---

[11]Chalmers 2011.

dispute. Indeed, I'll argue in chapter 4 that this claim of sufficiency and exclusion the only plausible way to understand Scanlon's argument.

Most of what I have to say about the buck-passing view focuses on the negative thesis. But the positive one (that goodness is a higher-order property) deserves some attention too, even if only to show why it's not more central to the dissertation. Philosophers seem to agree that the idea of analyzing goodness in terms of reasons has its source in Franz Brentano, who hoped to make human psychology and phenomenology the conceptual center of philosophy, ethics included. He writes,

> whoever loves and hates rightly, has his feelings adequately related to the object, i.e. the relation is appropriate, suitable, [and] corresponds suitably...[12]

This is the first expression of the *fitting attitude* theory of value. This view seeks to define or analyze value concepts in terms of a relation between ideal attitudes and the things that have value. In Brentano's version, to be good is to be worthy of love, and to be bad is to be worthy of hate. The view thus offers a conceptual analysis of values in terms of fittingness (or correctness, worthiness, etc.). Because fittingness seems to imply the existence of normative reasons to do the fitting thing, we appear to have an analysis in which goodness is defined in terms that are closely connected to reasons. Later versions of the fitting attitude approach from Henry Sidgwick,[13] A.C. Ewing,[14] Elizabeth Anderson,[15] and Justin D'Arms and Daniel Jacobson[16] have added detail and sophistication, but the core idea remains.

---

[12] Brentano 1889/1902, p. 70.
[13] Sidgwick 1907, pp. 110-111.
[14] Ewing 1948, ch. 5.
[15] Anderson 1993, p. 20.
[16] D'Arms and Jacobson 2000.

Yet however the fitting attitude tradition might have influenced Scanlon and the positive buck-passing thesis, it's far from obvious that they come to the same thing. This is because it's not clear whether fittingness is a value. One natural way to think about values is as ends—they're the aims of the various spheres of reasoning. In the realm of epistemic reasons, the goal is to believe well. For moral and practical reasons, it's to act rightly, well, or at least prudently. In aesthetics we aim to create beauty of some kind, in etiquette we want to be courteous, and so on for other kinds of reasoning. Fittingness lends itself nicely to this way of thinking: it's the end of practical reasoning in general, or perhaps of reasoning all-things-considered. If that's right, fittingness would be a kind of general value of which other, more familiar values are the species. We needn't take a stand on whether fittingness is a value, but there's a plausible case that it is, in which case the fitting attitude analysis and the buck-passing view would diverge in an important way. Whereas the buck-passer would account for goodness in terms of *non-value* properties that give reasons, the fitting attitude theorist accounts for it in terms of other *value* properties that give reasons.

None of this is to say that the fitting attitude analysis is inconsistent with buck-passing, but it might be consistent with other views too. And although there would be something odd about accepting the positive buck-passing thesis and thinking that values gave the best and most ultimate reasons, nothing about the positive thesis directly implies that those reasons come from value-makers either. As I'll show in chapter 4, the failure of the negative thesis undermines the case for the positive one, and that's why it's not the primary focus of the dissertation.

When we consider it as a whole, the buck-passing view looks attractive for much the same reason as the the fitting attitude analysis, namely because it offers a way to be a normative realist without the metaphysical excesses of mooreanism or platonism. Both the fitting attitude and buck-passing approaches tie goodness down to more concrete phenomena, namely attitudes and familiar properties, which is certainly an improvement. But metaphysical respectability isn't the only thing we want in a normative theory. It's also important to get the normative story right: the fundamental reasons identified by a theory should hold up to scrutiny too. Buck-passers (or at least those who adopt the value-maker solution) try to align these two kinds of explanation by having value-makers give both. And that can be a problem because normative and metaphysical explanation aren't the same thing.

To see this, consider a couple of examples that will feature prominently in later chapters. First is the old statue-and-clay pair, Lumpl and Goliath. The metaphysical story of Goliath the statue comes from Lumpl, the clay that composes it. But if we had to choose one to be the source of the reasons we have to, say, preserve the statue and put it on public display, Goliath seems to provide the better story. We should preserve it because it's Goliath—Lumpl is incidental. Here the metaphysical story differs from the normative one. Similar things are true of pain and the c-fiber firings it supervenes on (as the convenient fiction goes). One natural metaphysical explanation of pain is that it's grounded in the action of the c-fibers. But if we had to choose between the c-fibers and the pain state as the explanation of why we should help the afflicted, the pain is the better one. What matters is that the person is suffering, not that there's activity in some area of the brain.

These examples illustrate two things. One is the familiar point that semantic content makes a difference in explanation, and the second is that normative and metaphysical explanation come apart. Although the buck-passer would be happy to accept the examples because neither explicitly makes reference to values, that's intended. All the cases are meant to show is that we shouldn't move from the fact that *x is a metaphysical ground for y* to the claim that *x gives more ultimate reasons than y*. Buck-passers might well be right that in some sense the value-making properties and their relations are all we need to explain what reasons consist in. Unfortunately, good metaphysics doesn't always make good rational explanation—semantic content matters a lot for the latter, and much less for the former. We'll take up this point again in chapters 4 and 5.

2.3.2 Solution 2: values

Next is the view I advocate, that values give the best and most fundamental normative reasons we have. Call this the *value-first view* of reasons. I'll start by illustrating the view, then discuss some historical precedents for it, and consider some basic objections.

Start with the level of ultimate reasons. Because normative reasons are a kind of explanation, the view says that the best answer we can give for why someone ought to do something is that it would be good in some way. So for example, if I should shovel the snow from my driveway, the best explanation will be something like the fact that it's in my best interest to be able to get the car out. Or if I should take the dog for a walk, the best explanation will involve the benefit to her health. These examples cover the "best" half of our "best and most fundamental" slogan.

This scheme of explanation isn't limited to goodness of course. Badness works the same way. Where the fact that it's risky counts against investing all my money in one company, this is because it's imprudent to take that much of a gamble. Dissonance between two notes might give me a reason not to write them into the same song, and that reason comes from the fact that this case of dissonance takes away from the song's aesthetic value. Comparative value facts work too. If B is a better option than A, that can explain why the considerations favor choosing B.

That values both good and bad give ultimate reasons is compatible with allowing non-value considerations to give looser, everyday reasons. The view stops short of the claim that the only considerations that favor are considerations of value. It would be a needless cost to suggest that everyday claims about reasons are strictly false, like "the traffic is a reason to leave for the airport earlier than usual." This is an improvement over the buck-passing view, since buck-passers appear to be committed to denying the truth of any value-involving claims about reasons.

Thus we have explanations of why someone ought to do something that are passable but not as good as the ultimate reasons. At this level there are genuine reasons that appeal to value-making properties rather than values themselves. Since those reasons shouldn't be considered independent from the ones given by the relevant values, they must be dependent on them. So if the fact that the morning sky has suddenly gone dim gives me a reason to believe it's going to rain soon, that's because the arrival of dark clouds is good evidence for thinking it will rain. If the fact that giving money to Doctors Without Borders will lead to the eradication of polio in Southeast Asia counts in favor of my writing a check to them, its being a reason is explained by the

moral goodness of doing so. Ordinary considerations get their normative pull from values in this way—that's what it means to say that reasons come from values. These examples illustrate the "most fundamental" half of our slogan.

Some would conclude from the idea that values give the ultimate reasons that non-value facts must give derivative reasons—that is, derivative on the independent reasons values give. In one sense they are. They're derivative by not being the end of the explanation of why something is favored. But this is a fairly weak kind of derivativeness, since they're still closely related to ultimate explanation by supervenience. This contrasts with a more objectionable way a reason could be derivative: by being a mere place-holder for the genuine thing. To borrow Mark Schroeder's example, the fact that there's a reason to go into the living room is a reason to go, but it's wholly derived from the real reason, which is that your friends have planned a surprise party for you.[17] Non-value facts we appeal to as reasons in ordinary contexts don't depend on values in this way, so the reasons they give won't be derivative in the stronger sense. That wraps up the basic statement of the view. Now we'll turn to some history.

The eudaimonistic outlooks of Plato and Aristotole square nicely with the idea that values rather than value-makers give the ultimate reasons. Since they thought wisdom requires aiming at the good life in all things, everything we have reason to do will be for the sake of the good life. This comes out especially in Aristotle. He argues in Book 1, chapter 5 of Nicomachean Ethics (1095b15-1096a10) that people should choose subsidiary goods like pleasure, honor, and weath because they are conducive of the good life, not because they are what the good life actually is.[18] And he says later in Book 1 (1097b15)

---

[17]Schroeder 2011.

[18]See a similar point in Kraut 2014.

that the good is what makes things desirable or choiceworthy. These passages make a good case that Aristotle thought well-being was the fundamental normative explanation, even if he didn't employ the modern concept of a reason. They also suggest that Aristotle did not succumb to the temptation to align normative explanation with metaphysical explanation, as Plato and the fitting-attitude/buck-passing theorists seem to do. Whatever the metaphysical story for the ultimate good turns out to be—for Aristotle, this appears to be the virtuous activity of the soul (see 1097b-1098a20)—it won't be the normative one, since it's not the thing in virtue of which we ought to do everything else.

Because of Aristotle's place as the figurehead of the modern virtue ethics tradition, some may want to resist associating him with the idea that reasons come from values, since the modern virtue approach is often presented as an alternative to ethical approaches that focus on goodness or rightness, namely consequentialism and deontological views. We'll address the value-first view's similarities with consequentialism further on, but we can start to make it more palatable for modern-day virtue theorists by highlighting what makes virtue ethics an interesting methodological alternative to consequentialism and deontology: it dispenses with actions as the primary bearers of normative properties. Insofar as it's desirable to have virtues and dispositions at the heart of normative theory, the value-first view can accommodate that. Virtues are values too—values of character, intelligence, etc.—capable of generating moral reasons, epistemic reasons, and so on.

The modern inspiration for the idea that values give the *best* reasons is G.E. Moore's consequentialism, and most philosophers these days treat him as the ideal representative of the view. Moore famously claims in the *Principia*

that goodness is the one and only property that makes things right, which is evidence for thinking he would have called a consideration like $<\varphi$-ing would be good$>$ the best kind of moral reason to φ.[19] Moore didn't often make use of the reasons-talk that we do now, but there are a couple of passages where he mentions them. The first is from *Principia*.

> The only reason I can have for aiming at 'my own good' is that it is *good absolutely* that what I so call should belong to me...But if it is *good absolutely* that I should have it, then everyone else has as much reason for aiming at *my* having it, as I have myself.[20]

And from 1942:

> ... the fact that an action, which I could do, would produce *some* intrinsically good thing is always some reason (though far from a conclusive one) in favor of the hypothesis that I ought to do that action: or, in other words, that such a fact is always *favorably relevant* to the hypothesis that I ought to do the action in question.[21]

Taken together, these three statements make a strong case that (a) Moore thought goodness always gives the best moral reasons, and (b) he had something like the standard conception of a normative reason in mind. But as it was with Scanlon above, we need more to get us from this view to the idea that values give the best *and* most fundamental reasons, both for moral reasons and other normative reasons alike.

To get us closer to the thesis we're looking for, consider what Moore says near the beginning of *Principia*, chapter 6:

---

[19]1903, p. 18

[20]1903, p. 150

[21]1948, p. 565.

> ... I regard it as indubitable that Prof. Sidgwick was so far right
> ... that such mere existence of what is beautiful has value, so small
> as to be negligible, in comparison with that which attaches to the
> *consciousness* of beauty. This simple truth may, indeed, be said to
> be universally recognized. What has *not* been recognized is that it
> is the ultimate and fundamental truth of Moral Philosophy. That it
> is only for the sake of these things—in order that as much of them
> as possible may at some time exist—that anyone can be justified
> in performing any public or private duty; that they are the *raison
> d'être* of virtue; that it is they—these complex wholes *themselves*,
> and not any constituent or characteristic of them—that form the
> rational ultimate end of human action and the sole criterion of
> social progress: these appear to have been truths which have been
> generally overlooked.[22]

The idea that consciousness of value forms the ultimate end of human actions sounds much more like the target view. We already had the claim that values give the best reasons, and now we have the claim that they are thing in virtue of which all other things should be done. We do not, however, have any reason to think Moore was motivated by the rational competition problem, that he thought in terms of levels of reasons, or that he would have countenanced a wide range of values (although he did of course think there are many different things that are intrinsically valuable). So at the very least, what we have is a very close cousin of the value-first view of reasons.

So far, locating Moore in the present debate might seem pretty straightforward. But there's an objection to the standard interpretation of

---

[22]1903, pp. 237-8.

Moore, claiming he never thought that goodness or other values give reasons. Jonas Olson argues that Moore thought goodness was just the metaphysical ground of rightness, and that reasons come only from natural properties. This would be, in effect, the negative buck-passing thesis without the positive one.

The first premise of Olson's argument that Moore didn't think goodness gives reasons is that they genuinely don't. The second claims that Moore's early and later ethical views are consistent with this fact. This is a weak argument even if we grant the premises, since it relies on the principle of charity to go through, and charity alone isn't much of a reason to believe Moore thought this. Olson's rationale for the first premise is Scanlon's redundancy argument, which we can set aside for now.[23] We'll consider its merits at length further on in the dissertation, suffice it to say there's good reason to be suspicious. The second premise is more interesting for present purposes.

To show that Moore's views are consistent with the idea that goodness doesn't give reasons, Olson separates the outlook of the *Principia* from the one found in *Ethics* and later writings. His interpretation of the early Moore is that when something has the right natural properties, it has the further property of being good which "analytically implies" the existence of reasons, but doesn't give any.[24] The argument for this interpretation seems to be that there's a difference between giving a reason and being the analysis of a reason. So it's consistent for Moore to say that natural properties give the reasons, while analyzing those reasons in terms of goodness: x is a reason $=_{df}$ x is productive of goodness, or something similar.

Similar things go for the later Moore. Olson suggests that Moore is perfectly able to agree with the buck-passer that non-value properties give

---

[23]Olson 2006, pp. 526-527.
[24]Olson 2006, pp. 527-529.

all the reasons. All Moore would saying is that there being a reason is materially equivalent to there being goodness around.[25] On this reading, the only difference between the early and later Moore is that where the connection between goodness and reasons is analytic in the *Principia*, it's weakened to mere necessary correlation later on.

Now, Olson is surely right that if Moore restricted himself to just these views about the connection between goodness and reasons, he could also have denied that goodness gives reasons. But we already have plenty of evidence that he went a good deal further than that, all the way to thinking that values (or at least consciousness thereof) gave the best reasons.

Olson acknowledges that the first two passages from Moore we saw above—the ones that actually say goodness gives reasons—make it look like that's what he thought, but Olson urges us not to take them literally. Instead, he suggests, talk of goodness giving or being a reason *may* just be short-hand for saying that the natural properties goodness supervenes on give all the reasons, and we can give this short-hand because any talk about goodness implies that there are reasons around.[26] This would be a plausible reading of the passages only if there was independent textual evidence for thinking Moore thought goodness didn't give reasons, and Olson has none. In light of that, the text we have speaks in favor of believing Moore did think of goodness as a reason-giver, and also perhaps that he thought it gave the best and most ultimate reasons.

So much for Olson's objection to the standard interpretation of Moore. To close out this section, we'll look at some other common objections to Moore that might be raised against the value-first view.

---

[25]Olson 2006, p. 529.
[26]Olson 2006 p. 528 and p. 530.

Moore's general outlook is now long out of fashion. There are a handful of reasons for this, but fortunately the common objections to full-blooded mooreanism need not confront the value-first view, which is substantially weaker. The first group of these is against non-naturalism.

Mooreanism holds that goodness can't be analyzed, i.e. that it's a *sui generis* and simple property with no proper parts. This can be troubling for those who prefer a purely scientific view of the world, one free of irreducible normativity. Contemporary metaethical naturalists with this mindset often claim that the normative can be analyzed with identity: normative properties like *being good* are *a posteriori* identical to non-normative scientific properties, like *being conducive to human cooperation and survival*, etc. Although Moore himself would not have accepted this, it's entirely compatible with the value-first view of reasons. The varying kinds of goodness and badness might be identical to arrangements of non-normative properties, but give reasons just in virtue of their being values. Although I'll continue to refer to the non-normative properties values supervene on as value-makers or realizers and talk as if they're not identical, this shouldn't be taken to preclude identity between them. Identity is just a special case of supervenience and realization.

That said, the view does require that only values give the ultimate reasons, and some might worry that we can't have it both ways if we're identity naturalists. Suppose that values give the best reasons and are identical to arrangements of natural properties. Those arrangements would give the best reasons too, right? Not exactly. As we saw earlier, it's not the case that if x gives reasons and x is identical to y, then y gives reasons too. Think of pain and microphysical properties again: they might be identical, but if we can't understand the micro properties, then considerations featuring them cannot

be reasons, because considerations must be appreciable in some sense in order to be normative reasons. So contemporary non-naturalists should feel free to adopt the view, remembering of course that non-value properties can give everyday, non-ultimate reasons.

A second kind of objection focuses on epistemology. Moore's intuitionism has it that knowledge of moral facts is non-inferential and a priori. The view that values explain reasons isn't committed to anything as strong as that, so the many of the usual objections to intuitionism—e.g. that there is no such faculty of intuition, or that there's no way to tell if our intuitions are getting things right—can be safely put aside. But since the view is compatible with non-naturalism about values, there might remain a worry about how such abstract properties would be accessible if they're not identical to more familiar ones.

This point is structurally analogous to the Benacerraf Problem for platonism about numbers. According to that objection, it's hard to see how we could know anything about numbers if they were real, abstract, and mind-independent since they would have no causal connection to us.[27] Although the problem is *prima facie* compelling and interesting in its own right, there is an important difference between numbers and values that should prevent it being a concern for non-naturalism about values. Whereas platonic numbers aren't realized by more ontologically basic properties, values are. Because many of these value-making properties are easily accessible (like someone's experiencing pleasure, or the traffic's causing us to miss our flight) all we need is a belief that certain value-makers correlate with values, and we have inferential access to them.

---

[27]See Benacerraf 1973.

One might wonder why we need values to give reasons at all if their realizers are epistemically sufficient for us to know about the normative world. But just as we don't want to confuse our epistemology with our metaphysics, we don't want to confuse it with our normative explanation either. Although questions about justification are normative and thus part of the sphere of reasons, this is a worry about access, not justification. So it's really a metaphysical question rather than a normative one, and we've already seen that metaphysical explanation doesn't tell us much about the normative.

The last group of objections to consider are against consequentialism, and here we mix worries about Moore with more general concerns. Saying that values give the ultimate reasons sounds very much like a version of consequentialism. There are, of course, too many concerns about consequentialism to address here. Yet many of these can be dealt with by pointing out two ways in which the present view differs from classical consequentialism (CC).

First, some versions of CC place the explanatory burden on special kinds of values, in particular:

- Final values (things valuable for their own sake)

- Intrinsic values (things valuable just in virtue of their intrinsic properties)

- Absolute or unqualified values

The value-first view of reasons doesn't on its own require that any of these be given special explanatory significance. It's consistent with the idea that values give the ultimate reasons to allow that all kinds of values—moral, epistemic, prudential, aesthetic, instrumental, etc.—can be the end of rational

explanation. Thus the view is also free from commitment to value monism, unlike many versions of CC. That said, it is attractive to think that final values do enjoy a kind of priority—while instrumental goodness does some genuine explaining, the further down the chain we get from something valuable for its own sake, the less ultimate it is.

Second, the value-first view differs from CC because it makes no claims about weighting. That is, it need not be a maximizing view in the sense that what's most good gives the strongest reasons. The idea that reasons come from values lends itself naturally to maximizing, but for present purposes we're just considering the singular claim that value properties form a class, namely the class of things that give the best and most fundamental reasons. This appears to be logically compatible with non-maximizing. Leaving this question open has an additional benefit, which is that the view can allow for rational supererogation: things might be better without giving us more reasons to do them, and the attractiveness of supererogation could explain why.

So, is this consequentialism or not? Well, it's certainly not utilitarianism or mooreanism. And it's consistent with a range of deontological views, because it takes no stand on what the values actually are—remember, it's just a theory about the proper form of normative explanation. Properties like respecting the rights of others or complying with the categorical imperative are plausibly thought of as values too. We don't need to see the view as consequentialist, but to the extent that it is, it doesn't seem especially objectionable. Many, indeed most, moral views can be interpreted as kinds of consequentialism.[28]

One of the virtues of the idea that values give the ultimate reasons is that

---

[28]See Dreier 1993 for a defense of the idea that any moral view can be consequentialized, and Brown 2011 for a proof that there will be a small number of views that escape consequentialist assimilation.

it represents a middle way in non-naturalist realism between the dominant buck-passing tradition and mooreanism. It avoids the platonic excesses of Moore, while staying clear of the excessive reductionism in buck-passing (more on that in chapter 5) In the chapters that follow, I'll try to argue that the view is true and show that the primary motivation people have for denying it—Scanlon's redundancy argument—leads to serious trouble.

2.3.3 Solution 3: both

Finally, consider the idea that there is no unified solution to the rational competition problem: some ultimate reasons are given by values, and others by value-makers. It's not clear whether anyone has held this view (although both Jonathan Dancy and Roger Crisp have said things that sound like this).[29] There are two ways of developing this thought.

On one version, we deny that there's a unified solution, but we do so for principled reasons. There would be rules for determining when a value property is the source of a reason, and when the source is a value-making property. This is fine as far as it goes, but it's hard to see what those rules might be. Since there are no attempts to state such rules as far as I know, we can safely set this version of the view aside for now.

The other version does away with a unified solution, but offers no rules for fixing when one or the other is the ultimate reason. Rather, specific and potentially unique features of individual contexts determine what the most basic reason is. This kind of view might be motivated by a desire to respect the complexity and messiness of lived experience. That's a noble aim, but this

---

[29]In separate places, Dancy has said that values don't give reasons, and also that values are the "ground" of reasons. See Dancy's 2000, p. 164 for the first claim. See his 2002 for the second. Crisp claims that goodness sometimes gives reasons, and sometimes it doesn't, in his 2008.

version strikes me as a non-starter under present purposes. As we saw above, one of the primary challenges for realism about normative reasons is to show that they are well-behaved enough to be taken seriously. And since nearly everyone should be a realist of some kind, rejecting simplicity and order in favor of capturing the messiness of experience is to make realism much more susceptible to the challenges of anti-realism and skepticism. If we can't state an orderly theory of reasons, then the whole idea of normative reasons begins to look more like a story we tell ourselves and less like a genuine phenomenon that there are facts about. I write this in full knowledge that the kind of person tempted by this version of the view might think the order and simplicity of a unified solution creates a much more serious threat of skepticism because it alienates people from the reasons they have (as Bernard Williams or Michael Stocker might say).[30] I just disagree. In chapter 3 I'll develop a view about motivating reasons that enables us to keep the idea that people act well when their motivating reasons coincide with their normative ones, while still allowing that the thoughts that occur to us might not be as cold and calculated as philosophers' descriptions of those reasons often suggest. So at least some worries about alienation can be handled without resorting to this view.

That's all for the main portion of the chapter. If you have no qualms about accepting the standard conception of a normative reason, feel free to skip ahead to chapter 3. There are, however, philosophers who may be skeptical that *considerations counting in favor of actions or attitudes* is the best way to capture the phenomenon of reasons. These next few pages are for those people.

---

[30]See Stocker 1973, and Williams 1973.

2.4 Worries about the standard conception

The view that reasons are considerations counting in favor of actions or attitudes is widely adopted, but it would be nice to have an explanation of why it's the standard, and why it's safe to assume here as well. To close out the chapter, I'll address each component of the counting-in-favor relation as it relates to the questions under discussion here.

2.4.1 Favoring

Why say that considerations *count in favor* rather than something else? The standard answer is that favoring is conceptually equivalent to being a reason, so we can't do any better—they're foundational concepts of the normative realm and they can't be explained any further. As Scanlon puts it,

> I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to same idea: a consideration that counts in favor of it. "Counts in favor how?" one might ask. "By providing a reason for it" seems to be the only answer.[31]

Derek Parfit says similar things in his recent book.[32] For present purposes, I think it's safe to assume this view or something close to it. The concept of counting in favor does appear to be pretty foundational. It's worth mentioning that there could be reasonable doubts about the idea that favoring and reasons can't be explained any further, though. If we want to know how something counts in favor, "by providing a reason" doesn't seem to be all we can say. "By showing it's worth doing" or "by explaining how it would be good" also

---

[31]1998, p. 17.
[32]See the first chapter of his 2011.

seem to be reasonably informative answers, but we'll leave that question aside here. This dissertation is about which things give our best reasons, not the conceptual analysis of reasons. So nothing to come turns on whether the standard conception is right about this.

2.4.2 Considerations

Now on to what the favoring relation relates. The standard conception says that considerations favor, and the usual idea is that considerations are facts or true propositions. Why *facts or true propositions*? This is best thought of as a question about why considerations need to have a close connection to truth and falsity, as facts and propositions do, rather than being things further removed from truth: a desire for ice cream or the property of being worse than the alternatives. One thought is that we want a near relation to truth because we're talking about genuine favoring—normative reasons are constrained by the way things actually are. While this general thought seems right, doing it justice doesn't seem to require facts or propositions. So while truth might be an easy way to get at the actual, things like properties, desires, etc. can also deliver it. This leaves open the possibility that considerations involve more than facts or propositions. As we noted earlier, ordinary language suggests we're comfortable with that. Think of all the times we say things like "the market value of your house gives you a good reason to wait a few years to sell," or "the good that Oxfam does gives you a reason to donate." There might be reasonable ways to paraphrase this kind of talk away, but these attributions of reasons are not obviously false. Moreover, since reasons stand for normative powers, attributing those powers to something like propositions might play back into the hands of skeptics, nihilists, and deflationists. Many deny that

propositions exist, or are at least suspicious of them, and so any theory giving them powers may look dubious.

The lesson to draw from all this is that the standard view of considerations as facts or propositions is just a model. It does not settle any deep metaphysical questions about what they are—like any model, it just needs to approximate its target. The true propositions we assign as considerations supervene on the less abstract alternatives I mentioned above (properties, desires, etc.), so we can expect they'll be close enough. From now on, I'll talk of considerations as facts, by which I mean they are true propositions. I do this to follow established usage, and as we'll see further on, there are good reasons for sticking with this usage. Yet for present purposes we needn't take a stand on the metaphysics of considerations. The same goes for the relation of counting in favor and what considerations count in favor of. We'll be concerned instead with the normative explanation of these things.

2.4.3 Actions and attitudes

The last element to ask about are the things that considerations count in favor of. The standard conception says that these are actions or attitudes, like driving a friend to the hospital or intending to walk the dog. Why these? Scanlon answers that it's because these are things we can intentionally do and be responsible for—in Scanlon's terms, they're the things we're open to judgement for.[33] That is, actions and attitudes are the focus of reasons because they're they only things that can be regulated by normativity. That we have control in some sense over these and not other things is the explanation of why they're governed by norms. Notice that this answer to the question is

---

[33] 1998, pp. 21-22.

effectively the same as the popular view that *ought* implies *can.* I'll say more about the Ought Implies Can principle in chapters 3 and 7.

I agree with the basics of Scanlon's explanation, but the way it's pitched leaves it open to controversy. Those who reject the idea of moral or normative responsibility will balk at his way of putting things. Yet Scanlon's idea about being open to judgement can help here. Although judgement sometimes involves holding people responsible, it need not. It's also a way of evaluating: A did well, but B acted badly, and so on. And we can make these evaluative assessments without holding that someone is deserving of praise or blame in the way that incompatibilist moral responsibility requires. They are also appropriate even if determinism is true.[34] So the better explanation of why actions and attitudes are what considerations favor is that those are the things we can be appropriately evaluated for. Some attitudes are the right ones to have, some beliefs are simply unjustified, and some actions are imprudent.

Evaluable actions and attitudes may not be the whole story of what reasons count in favor of, though. Think of people with lots of bias. They might be evaluable not just for questionable actions and attitudes, but also for failing to do anything about their disposition towards them. And although actions and attitudes are essential to any dispositions thereto, you might think that what's being evaluated with dispositions may be different such that the judgement of actions and attitudes may not account on its own for what we think of the related disposition. Take a person who forms a racist attitude but doesn't have the disposition to do so, and then think of a person who forms the very same racist attitude and has a long-standing habit of doing so. It doesn't look unreasonable to think that the same reasons tell against the attitude that

---

[34]Here I am parroting a famous point from Gary Watson. See Watson 1996.

these two people share, but judge that there are additional reasons against having the disposition. This is tied to the virtue-theoretic idea that there are moral evaluations of actions, but there are also moral evaluations of character or people. Maybe there are reasons to instantiate properties too.

I think all of this is plausible, but for ease of presentation we'll stick to the standard idea of actions and attitudes as the objects of reasons in the rest of the dissertation. If it turns out these other things should be added to our model of reasons, then everything I have to say about the source of normative reasons will hold for them too: if there are reasons for having dispositions or properties, values give the best and most ultimate reasons for having them.

Chapter 3

Ought Implies Can, Can Implies Value

Like a lot of people, I believe in something in the neighborhood of the views known as *Ought Implies Can* and the *Guise of the Good*. This chapter does something ambitious with these ideas. Versions of them will be put to use as premises in an argument that values give the best and most ultimate reasons we have—that is, the *value-first view* of reasons. Both premises are controversial, but I'll argue that they can avoid common objections and still work to imply the conclusion we need.

In broad strokes, the argument goes like this. If *ought* implies *can*, then because normative reasons place us under *prima facie* obligations, the existence of an ultimate reason implies that it's possible to comply with it. That's because practical wisdom requires not only doing the proper thing, it requires doing it for the right reason. If wisdom means doing thing x because of consideration y, then truly complying with rationality means doing x because of y. To do anything less would be to do things that we don't fully understand.

And if the Guise of the Good is true, then the ultimate explanation of our motives must be that we judge our actions to be worth doing—that is, valuable in some way. The explanation of our motivations when we act well will line up with our ultimate normative reasons, because that's what practical wisdom amounts to. Since the space of ultimate reasons is constrained by what's motivationally possible, and what's motivationally possible is limited to considerations of value, all the ultimate reasons are considerations of value.

Normative reasons contrast with motivating reasons, which are also central to the argument. Motivating reasons involve a consideration and an object too. But in the case of motivating reasons, the consideration explains an agent's motivation towards the object rather than favoring it. So when I'm moved to stop writing my dissertation and go play some tennis, the explanation might be that I felt it would be better to get some exercise and fresh air than to stay inside all day. In that case there's a consideration, <that it would be better>, which explains my motivation towards the object of playing some tennis. It's helpful to describe motivating considerations in this explanatory sense because there does not seem to be a clear folk theory of motivation—people describe their motives as beliefs, desires, and emotions. And while all of these states can take propositions as their intentional objects, it might seem phenomenologically wrong to suggest that we are always act with value propositions in mind.[1] Describing motivating considerations as explaining our motives rather than being them avoids this worry.

Motivating considerations also face a rational competition problem when understood in this way. When I get up to play tennis, we can give a complete explanation of my thinking in terms of value: <that it would be better for

---

[1]Some people think emotions aren't propositional attitudes. We'll leave that question aside here, as it makes little difference to the argument of the chapter.

my state of mind>. But we can also give a complete account in value-making terms: <that it will reduce my stress hormone levels>. Counting these as independent motivations would lead to a double-counting problem, just as we saw with favoring considerations. And just before, the the best way to dissolve the competition is to make one explanation more fundamental than the other. Thus we have explanations of people's motives that are more ultimate than others.

Considerations are best thought of as facts or propositions, depending on what role the consideration is playing. If a consideration is involved with a normative reason then it will be a fact, understood as a true proposition. False propositions don't objectively favor anything. But when the consideration is part of a motivating reason, it need not be true—we're often motivated by mistaken ideas. Thus a consideration of value is a proposition about something's having value. More specifically, a proposition is a consideration of value just in case it's a proposition to the effect that something satisfies a value predicate, like 'good', 'bad', 'better', 'worse', etc. So that's why <that it would be better to get outside> is a consideration of value.

Now that the main concepts are a bit more precise, here's what the argument looks like. Let's interpret Ought Implies Can as saying that every ultimate reason is a possible ultimate motivating consideration. And we'll take the Guise of the Good to say that every possible ultimate motivating consideration is a consideration of value. Together, these premises imply that every ultimate reason is a consideration of value. They bar non-value considerations from being ultimate reasons because they don't play the correct motivational role.

This argument assumes (a) that there really normative reasons, and (b)

that some are more ultimate than others. Recall from earlier that (a) can be interpreted as minimal truth-realism about reasons, the view that at least some sentences asserting the existence of a reason are true. And because we could never have a reason in favor of denying this view, it seems safe to accept. Assumption (b) is a consequence of the rational competition problem. Since we have everyday reasons that compete with one another, the best move is to have one kind of reason be the normative foundation for the others, the things in virtue of which the other reasons hold. The alternative would mean denying that some of the competing reasons are genuine, but that would diverge significantly from general linguistic practice.

Before we consider the premises directly, it's worth saying a bit more about what the argument shows. The premises are meant to show that values give the best and most ultimate reasons. Calling this the value-first view is an improvement over two other names that are sometimes used for the idea that values give reasons: mooreanism and teleology. We saw in chapter 2 why the view isn't the same as Moore's, but we haven't yet discussed teleology.

On its face, the value-first view may look a lot like what Scanlon has called "teleological" conceptions of value, but it's actually quite different. In Scanlon's terms, teleological conceptions of value have two distinguishing features that the value-first view is silent about.

Teleological views take states of affairs to be the sole bearers of intrinsic value and direct us to bring about the best possible states of affairs.[2] By contrast, the value-first view makes no claim at all about what things have intrinsic value. It's a view about what sorts of propositions are eligible to favor. It also says nothing about whether or not we ought to bring about

---

[2]Scanlon 1998, pp. 79-80.

what's best—in fact, it says nothing at all about how strong any reasons are. Again, it's only a view about eligibility, not what the balance of considerations actually favors. This is also why the value-first view is not a version of classical consequentialism, as we saw in chapter 2.

Before we move on, it's also worth revisiting what all this value-talk means. We often talk about value generally, in terms of unqualified goodness and badness. But this is best thought of as a generic way of referring to more determinate kinds of value, like moral goodness, epistemic goodness, prudential badness, etc. These more determinate values should map nicely on to the different kinds of normative reasons that there are: moral, epistemic, prudential, and others. These determinate values often have more vivid labels. Moral value can be described as rightness, epistemic goodness is warrant or justification, and aesthetic goodness is beauty, and so on. I don't have much of an argument for this claim, but recall the view of values laid out in previous chapter. Outputs of standards of evaluation for actions and attitudes often involve terms like rightness of justification. Moral standards tell us whether an act was right or wrong, epistemic standards tell us whether a belief is justified, standards of etiquette tell us whether an act was polite, and so on. Standards of evaluation could be understood as giving verdicts only in explicitly evaluative language, which is a possibility the value-first view could allow, but that's just not how we talk.

I'll often speak of values using these other descriptors, but for those who tend to think of justification and rightness (or whatever) are not kinds of goodness but rather good-makers, let me re-emphasize that the view takes justification to count in favor of things only in virtue of being a kind of goodness. This point is important because it marks an important

difference between the value-first and buck-passing views. On that view, every consideration that favors is a non-value consideration about good-making properties. If the view allowed that considerations of justification, rightness, or beauty counted in favor intrinsically rather than in virtue of being kinds of goodness, it would be in danger of collapsing back into buck-passing.

Here's the plan for the rest of the chapter. We'll look at both premises in detail, considering basic rationales for each. As they are both controversial in their own way, we'll address some common objections to both. This will require making some adjustments to the original ideas behind the premises, but I'll argue that there are plausible versions of both Ought Implies Can and the Guise of the Good that avoid the objections and can still do the work the argument requires.

3.1 Motivational Ought Implies Can

Classically, Ought Implies Can is the doctrine that if you're obligated to do something then it must be possible for you to accomplish it. The classical version focuses on moral reasons, but it's not hard to see how a similar thought would be true of all normative reasons—which carry their own kind of obligation—and not just moral ones. Our interest is with ultimate reasons, so the principle will be framed around them primarily. Interpreted according to the structure we get from the standard conception of reasons, Ought Implies Can (for ultimate reasons) can be read in two different ways. On the one hand is a thesis about the objects of reasons and on the other is a thesis about the considerations that give those reasons. The object-oriented version says:

If ultimate consideration c favors some object o for an agent S,

  then S can bring about o.

And the consideration-oriented version says:

  If ultimate consideration c favors some object o for an agent S,
  then c can be S's ultimate motivating consideration for o.

The object-oriented version is probably closer to the classical version of Ought Implies Can, and it might well be true. We'll leave that question aside here. To the extent that the basic thought behind Ought Implies Can is true of normative reasons, the object-oriented thesis is not the whole story. I believe the consideration-oriented version is also true, as it appears we need it to make sense of normative reasons as a philosophical concept. Let's call this version of the doctrine *Motivational Ought Implies Can*, or MOIC.

The primary reason for accepting MOIC is that if it's not true, full practical wisdom seems unattainable. The denial of MOIC would mean there are some ultimate reasons that can't be acted on, which means we cannot always act for the right reasons, thus violating the idea that ethics is essentially action-guiding. But normative reasons and the theories about them are in the end part of the philosophical project of ethics. The most universal description of ethics we can give is that it's an attempt to answer the question, "what should I do?" So ethics and the normative phenomena that populate it really are essentially action-guiding, and that's reason to accept MOIC.

As we saw in chapter 1, acting rationally requires an understanding of why the objects we are intending to bring about are favored—that is, why we ought to do what we ought to do. For example, suppose I need to take the the 9:27 bus to campus to avoid keeping my class waiting. I've acted in accordance with my reasons if I'm motivated to take the 9:27, say, by a desire not to

be impolite to the class. But I haven't acted in accordance with reason, and certainly not as well, if I'm motivated to take the 9:27 instead out of a mistaken belief that it's the only way to avoid nuclear war with the alien invasion force waiting just beyond the moon. I may have complied with the object of my normative reason under the latter motivation, but not the consideration that gives it. The same goes for ultimate reasons: acting in accordance with one's best reasons means acting on the considerations that give them.

This is meant to be a fairly weak requirement on normative reasons. It doesn't require that any ultimate reasons are actual motivating considerations, and it doesn't require that the objects of those reasons be possible. It requires only that every ultimate reason be a possible ultimate motivating consideration, in the explanatory sense I outlined above.

MOIC is a modal principle, dependent on some sense of possibility. The best candidate is psychological possibility, since no consideration can motivate unless it's psychologically possible to appreciate. Yet psychological possibility resists easy definition. Let's try to illustrate it with a few examples:

- Vividly imagining a cube that's both entirely red and entirely blue is psychologically impossible, but imagining that Elvis and Johnny Cash were actually the same person is possible, generally speaking.

- Understanding what it's like to navigate the oceans by electroception is psychologically impossible, but understanding the axiom of choice is possible, generally speaking, even for those who have never heard of set theory.

- Appreciating the full causal chain between a butterfly flapping its wings in Syracuse and next week's weather forecast in Paris is psychologically

impossible, but appreciating the role of unconscious bias in hiring decisions is possible, generally speaking, even for those who don't believe that there is such a thing as unconscious bias.

Even though this way thinking about psychological possibility is permissive about what's possible, claims of psychological possibility can only be made generally, and not universally. Accidents of age, injury, disability, etc., make it so that very little if anything is universally psychologically possible. Similar things may be true of impossibility.

Given that notion of psychological possibility, MOIC looks consistent with plausible versions of both internalism and externalism about motivating reasons. Roughly, internalism is the claim that normative reasons do motivate people under the right conditions—or, that favoring considerations are also motivating considerations under the right conditions—and externalism is the denial of this claim. Internalism about ultimate reasons would look a lot like MOIC, but they're importantly different. Because our notion of psychological possibility is so weak, there will be considerations that are possibly motivating even though there are no circumstances under which they actually do or would motivate the person in question. After all, some people are extremely stubborn, or in the grip of a strong bias. But as long as it's psychologically possible for such people to think differently (be a little less stubborn, etc.) to the point where they can be motivated by the consideration, many non-motivating considerations will turn out to be possible motivators.

So much for the rationale behind the premise. Now we'll consider some objections, with the goal of showing that MOIC can avoid them. Classical Ought Implies Can is seriously controversial, and problems with the classical view might be thought to carry over to MOIC. Addressing all of these concerns

would require a dissertation unto itself, and so we'll confine ourselves to three of the most popular.

The strongest objection to consider comes from *Hume's Guillotine*, the principle that you can't derive an *ought* from an *is*. Many people have thought the Guillotine is incompatible with classical Ought Implies Can. The idea is that if obligation implies ability, then inability implies a lack of obligation. Since the claim that something can't be done looks like a straightforward *is* and the claim that it's not the case that it ought to be done looks like a straightforward *ought*, the two doctrines appear inconsistent with one another.

Now, MOIC does imply that if a consideration isn't motivationally possible it cannot be an ultimate reason, and therefore it generates no obligations. But MOIC says this not as a matter of pure logic, but rather as a substantive and controversial view about what normative reasons do. As we saw, the motivation for believing MOIC has to do with practical wisdom and the basic project of ethics. By contrast, Hume's Guillotine is a view about what can be done with logic alone. As a matter of entailment, in order to validly infer *not an ultimate reason* from *not motivationally possible* we would need to include MOIC as a premise, and so a rational *ought* would appear in the premises. So far so good.

But here it's important to distinguish entailment from valid implication more generally. When premises entail a conclusion, the inference is valid no matter what interpretation we give to the non-logical vocabulary. So we can't move from *not motivationally possible* to *not an ultimate reason* under entailment even if MOIC is analytically true. But under valid implication we could, provided MOIC were in fact analytic. So even if it's not possible to generate counterexamples to Hume's Guillotine using MOIC under entailment,

it might still be possible to get an is-to-ought inference under valid implication.

There are two ways out of this problem. Either (a) show that MOIC is not an analytic or conceptual truth, or (b) find a good reason to reject Hume's Guillotine. I'm not optimistic about the prospects of (a), since I tend to think that MOIC might be a kind of conceptual truth, so (b) seems like the better option. But dismissing a principle so appealing and entrenched takes a lot of time, and would thoroughly derail the current argument. So I'll argue against the Guillotine in chapter 6, where I'll try to show that no available version of the principle is both true and adequate, and that giving it up has little cost because it can be and replaced with the moorean view that descriptive terms are never synonymous with normative ones. This view presents no trouble for either classical Ought Implies Can or MOIC. Even if they are conceptual truths, it won't be because 'ought' means the same as 'can'.

Then there's an objection from cases in which one becomes unable to do what one ought to, but the obligation remains all the same.[3] Suppose it's my duty to keep order in the classroom but that I'm unable to accomplish this because the students are just too unruly or the lectern microphone isn't working, and so on. These facts don't free me from my obligation, goes the idea. Thus I ought to keep order in the classroom, even though I can't. Cases like this would be counterexamples to the object-oriented version of Ought Implies Can, if the interpretation of the examples is correct. For my part, I'm unsure if it is. But for the sake of argument, let's suppose the interpretation is right—obligation can persist through periods of inability.

As the consideration-oriented version of Ought Implies Can, MOIC can make sense of that idea. In our present terminology, saying that I'm under

---

[3]See Stocker 1971, pp. 311-312; see also Howard-Snyder 2006, p. 235.

an obligation to keep order is to say that there's an ultimate consideration that favors keeping order. Suppose for the sake of argument that it's a consideration of value, like <it's the right thing to do>. According to MOIC, this consideration cannot be an ultimate reason if it's psychologically impossible for me to be motivated by it. But there's nothing in the case suggesting that it can't be a motivating consideration for me. I certainly can be motivated by it's being the right thing to do, because (let's suppose) I have already been motivated by it when I originally tried to establish order.

While MOIC it's compatible with there continuing to be reasons even when their objects may not be possible, it's also compatible with the absence of reasons in such cases. In other words, it's compatible with the object-oriented version of Ought Implies Can. MOIC is not a biconditional—just because one is capable of being motivated by a consideration doesn't mean that it's a genuine normative reason.

The second objection comes from what Michael Stocker calls "culpable inability." Cases of culpable inability are ones in which someone attempts to escape obligation by rendering himself incapable of performing it. A father who has promised to attend his child's recital, but decides not to go because he can't stand the sound of kids playing violins, might seek to free himself from his obligation by getting on a plane to Honolulu so that he'll be in the air while it's going on.[4] Or a teenager might shoot herself in the foot to avoid fulfilling her promise to go on walks with her grandmother, who she finds annoying. Clearly this is not behavior to be encouraged, and definitely not something that a good ethical principle should license. MOIC can handle cases like these in the same basic way as before: it's still psychologically possible for the father

---

[4]c.f. Stocker 1971, p. 314.

and the granddaughter to treat the consideration that <it's morally wrong to break the promise> as a motivator, even though they won't. So MOIC can allow that both still have good reasons to do the things they're trying to avoid.

But there's another kind of culpable inability case where this response won't work for MOIC. In these cases, people seek to avoid their obligations by altering their psychology, rather than rendering the objects physically impossible. Thus Jones might seek to avoid looking after his ailing spouse by hiring some unsavory characters to cause him irreversible brain damage, making it psychologically impossible to play the role of a caring spouse. Again, there's something seriously wrong with such behavior. But this time MOIC has to say that once the injury occurs, there are no considerations for Jones that favor caring for his spouse because his injury leaves him psychologically unable to take his wedding vows into consideration.

I think it's alright to bite the bullet here and agree that Jones has succeeded in getting rid of the normative reason for looking after his spouse. After all, injury is one of the things that interferes with psychological possibility. But biting the bullet means explaining away the objectionable features of the example. We need an account of why Jones's behavior is wrong, and here's a natural one: before he injures himself, there are many considerations that Jones can appreciate which give him reasons not to do it. There are moral considerations towards his friends, family, and spouse, as well as prudential considerations concerning the goods that the injury will deprive him of. His disregard for those reasons seems to enough to capture the discomfort we would feel about cases like this. Thus I think MOIC can also get around the objection from culpable inability.

So much for the first premise. Altering the focus of classical Ought Implies Can from the objects of our reasons to the considerations that give them offers a promising way to handle the usual objections. Now we'll turn from the idea that every ultimate reason is a possible motivating consideration to our second premise, which gets us from MOIC to the value-first view of reasons by way of the idea that the best and most fundamental explanation of our motivations is that we're moved by judgments of value.

3.2   The Guise of the Valuable

Going at least as far back as Plato, philosophers have been attracted to the idea that desiring something requires thinking of it as being good in some way. This thesis is known as the Guise of the Good. Much of the time it seems obvious and phenomenologically right. It can be difficult to imagine wanting something but not finding it good. The list of people who have endorsed versions of Guise of the Good includes Aristotle, the Scholastics, Kant, Anscombe, and Davidson.[5]

Arguments for the Guise of the Good often proceed from the idea that the view is necessary to make sense of intentional action.[6] Acting toward some desired end without thinking of that end as having some sort of value appears to be unintelligible. To borrow an example from Anscombe, suppose Smith is trying to acquire a saucer of mud but has no explanation why. If we ask her what she wants it for, all she can say is there isn't any end she wants it for, or any respect in which it's desirable—she simply wants it, full-stop.

---

[5]See Plato's *Meno* 77a-87c and *Gorgias* 466a-468e, the opening lines of Aristotle's *Nicomachean Ethics*, Aquinas's *Summa Theologica*, Anscombe 1957/2000, and Davidson 1980.

[6]See Tenenbaum 2013.

Smith's behavior would be pure nonsense, says Anscombe.[7]   Showing how an agent regards some action as good is just what intentional explanation amounts to, and so the Guise of the Good is a straightforward consequence of our explanatory commitments.

As with classical Ought Implies Can, the controversy surrounding the Guise of the Good is proportioned to its intuitiveness and popularity. Most objections fall into two classes. Some are concerned about whether goodness *per se* is always part of the content of our desires, and others are concerned with the possibility that motivation doesn't require any thoughts about value at all. We'll get to those objections shortly, but first consider an alteration of the principle that helps us get from MOIC to the value-first view.

Since this is meant to be an argument about ultimate reasons, it's unnecessary to insist that only value-considerations can motivate, just as it's unnecessary to say that only value-considerations give normative reasons. Other considerations can and do lead us to act, but only as part of a deeper motivational story. As we've seen, given that acting well means aligning one's motivating reasons with one's normative reasons, considerations of both kinds should have the same structure. That is, where normative reasons can be split into ultimate and non-ultimate reasons, so too can our motivating considerations. Thus the argument needs a principle framed around ultimate motivating considerations. Call this version of the thesis the *Guise of the Valuable*, or GOV:

(GOV): Every possible ultimate motivating consideration is a consideration of value.

---

[7] 1957/2000, pp. 70-71.

Notice the two changes from classical Guise of the Good. One is the change from desires to ultimate motivating considerations we've just been discussing, and the other is a move from goodness *per se* to considerations of value.

Recall that we're thinking of motivating considerations in the explanatory way. Ultimate motivating considerations are the most basic explanations of a person's action, and that's perfectly consistent with desires being the real psychological motivators. So the focus on desire in classical Guise of the Good can preserved in GOV (if that's important to you). The change to considerations of value is needed because the value-first view takes values of many different kinds to give reasons—like wrongness or justification—not just considerations of the form $<x$ is good$>$.

Now let's see how GOV handles the two common kinds of objection to Guise of the Good. The first kind suggests that people sometimes genuinely desire the bad, so classical Guise of the Good must be false. Both Michael Stocker and David Velleman pursue this line of reasoning.[8] Velleman gives the illustrative example of Satan in John Milton's *Paradise Lost*, who upon defeat resolves to start desiring the bad and avoiding the good, and argues that it cannot be that Satan mistakenly thinks of bad things as good. He correctly judges the bad as bad and chooses it under that guise.[9]

Since GOV is significantly weaker than classical Guise of the Good, it can handle these cases with ease. Badness is a value just as much as goodness is, and so the view is consistent with the notion that people can genuinely desire or be otherwise motivated towards the bad. Yet allowing badness to motivate

---

[8]See Stocker 1979, and Velleman 1992.

[9]Velleman 1992, p. 18.

may seem to conflict with the basic motivation for the Guise of the Good we saw above. The original rationale was that any explanation of intentional action requires showing how the agent saw his action as good in some way. You might think that if seeing an action as bad is also sufficient to explain, then it's intelligible for people to be motivated by *anything*, even evaluatively neutral saucers of mud.

There are at least two ways out of this worry. One is to update the basic argument for Guise of the Good. In Anscombe's example, it's natural to think of goodness as doing the work in intentional explanations, and this does seem to be the standard interpetation. But let's look at what she actually says. Imagine again that Smith has told Jones about her quest for a saucer of mud. Here's Anscombe:

> Would he [Jones] not try to find out in what aspect the object desired is desirable? ... Now if the reply is: "Philosophers have taught us that anything can be the object of desire" ... then this is fair nonsense.[10]

This passage makes it look like desirability is what really does the work in explanations of intentional action. This alone won't settle what things are desirable—if Stocker and Velleman's examples are right, then badness is desirable too, and is thus capable of motivating. To extend this conclusion to neutral things, we would need a successful explanation where no value appears at all. This sort of example constitutes the other kind of objection to classical Guise of the Good, which we'll get to shortly.

The second response is to highlight the role of ultimate motivating considerations. It could be that Milton's Satan is moved by evil *qua* evil,

---

[10]1957/2000, pp. 70-71.

but this may not be the end of the story. Someone can want the bad as bad instrumentally as a way of satisfying a desire to frustrate the aims of one's enemy, which would be a kind of goodness. On this telling of the example, Satan would genuinely be moved to badness, but non-fundamentally. His more ultimate motivation would still be given by positive value. This is, of course, just another version of the usual response to cases like this, but the introduction of ultimate and non-ultimate reasons gives us a way to preserve a sense in which atypical agents like Satan are genuinely motivated by badness, which other versions may have a hard time doing. Thus GOV need not be understood as the claim that ultimate motivating considerations are sometimes given by badness, even though it's consistent with it.

The second group of objections to the classical Guise of the Good comes from examples purporting to show that thinking of value isn't required at all for motivation. One such example comes from Jonathan Dancy. Suppose that Jonathan has promised his children that if they do their homework on time, he will tie his right shoe before his left on even-numbered days, and his left shoe before his right on odd-numbered days. Dancy says that he'll be motivated to tie his shoes in that way (provided the kids are doing their homework), even though he sees absolutely no value in it.[11] Another example comes from Joseph Raz, who describes a miner who is motivated to refuse a buy-out offer from managers that are intent on shutting down the mine. Asked if he thinks his refusal offers any hope of keeping the mine open, the miner will say no—it's just the right thing to do on principle.[12]

If these cases are right, then both Guise of the Good and GOV are false because motivation can occur without any consideration of value. I think

---

[11]Dancy 2000, p. 168.
[12]Raz 2010, pp. 112-113.

these are powerful objections to classical Guise of the Good, but the key to responding on behalf of GOV is there in the miner's answer. The miner says refusing the buy-out is the right thing to do, and presumably Dancy would say that following through with his promise is the right thing to do. While it might be that neither sees goodness *per se* nor prudential value in what they're doing, they do see moral value in it, insofar as they think they're doing the right thing. Because rightness is also a value, neither of these cases is a problem for GOV. I suspect the thought behind this kind of objection is aimed at thinking of things as being good or bad, without qualification. GOV allows for people to be motivated by the more determinate, qualified forms of value, and so it's not a target for this kind of criticism in the way that classical Guise of the Good is. To get an example that truly challenges GOV, we need a case of motivation in which no values of any kind could be playing an ultimate motivating role. To my knowledge, none of the critics of the Guise of the Good have offered any, and cases like Anscombe's saucer of mud suggest there won't be any.

That wraps up the argument for the value-first view: all the ultimate normative reasons are possible ultimate motivating resaons, and all the ultimate motivating reasons are considerations of value. Therefore, every ultimate normative reasons is a consideration of value. Or to put in a slogan, *ought* implies *can* and *can* implies value.

Chapter 4

Rational Redundancy & Rational Exclusion

Now that we've seen the positive case for the value-first view, we turn to the primary—and perhaps only—argument against it: Scanlon's redundancy argument. This chapter is about understanding how the argument works. I'll show that it has to be seen as a kind of exclusion argument if it's going to work at all. Properly understood, it alleges competition between goodness and good-makers for giving the same reason and concludes that the good-makers win out. This might seem rather obvious, but nailing down the terms of the argument is essential for the considerations I'll give against it in chapter 5.

Many philosophers seem to have found the redundancy argument persuasive, including many who don't also accept Scanlon's buck-passing theory of value. This is likely why so much of what's been written about the buck-passing view focuses on the Wrong Kind of Reasons problem and ignores the redundancy argument entirely. I certainly agree that there's something intuitively compelling about the threat of redundancy. If some

instance of goodness gave an independent reason from the reasons given by its good-makers, that seems like one reason too many, at least at first glance. That's how the rational competition problem arises in the first place. But few appear to recognize that we've heard this story before.

No so long ago, causal exclusion arguments against non-reductive physicalism about the mental were also quite popular. Roughly speaking, the worry was that because non-reductive physicalists believe mental properties are realized by—but non-identical to—physical properties, and that the causal powers of the mental supervene on the causal powers of the physical, their view leaves no causal work for the mental properties to do.[1] Take some mental state, which has physical properties P and mental properties M. The claim of the exclusion argument is that because any causal powers of the M would be had in virtue of the P's causal powers, the P must be causally sufficient. But if the P are sufficient on their own, then the M have nothing to add. If they did have something to add then there would be widespread causal overdetermination, which is tremendously unlikely and difficult to accept. Thus the reductionists concluded that if non-reductivism were true, nothing we do is caused by our mental properties.[2] Interestingly, this amounts to an exclusion objection against motivating reasons as well. From the beginning, non-reductivists like Davidson were concerned to make sense of how we act in virtue of higher-level mental states like beliefs, decisions, and desires.[3] If these things don't cause us to act, it's safe to assume they're not motivating us.

Now compare Scanlon's argument:

---

[1] For a nice explanation of the causal exclusion argument, see Fodor 1989.

[2] See especially Kim 1992.

[3] See for instance the opening paragraphs of his 1970b.

> First, when I consider particular cases it seems that these reasons are
> provided by the natural properties that make a thing good or valuable.
> So, for example, the fact that a resort is pleasant is a reason to visit it
> or recommend it to a friend, and the fact that a discovery sheds light
> on the causes of cancer is a reason to applaud it and to support further
> research of that kind. These natural properties provide a complete
> explanation of the reasons we have for reacting in these ways to things
> that are good or valuable. It is not clear what further work could be
> done by special reason-providing properties of value, and even less clear
> how these properties could provide reasons.[4]

The situation for normative reasons is structurally the same as the one that
sets up non-reductive physicalism for exclusion worries. We have realization of
value by value-makers, and we have supervenience of reason-giving powers. If
the natural properties of the resort had been different (if it had been even *more*
pleasant, say), then the reason-giving powers of the resort's goodness would
be different too (there presumably would be even *more* reason to go). Finally,
we also have basic commitment not multiply reasons needlessly. Scanlon tells
us that value-makers are already sufficient to account for all the reasons there
are, so given the basic facts above, he concludes that value is excluded: none
of the reasons we have are given by something's being good.

Now despite its intuitive appeal, the causal exclusion objection was not the
end of the story for non-reductive physicalism about the mental. The trouble
with exclusion arguments is that they exhibit the following weakness. They
are held hostage to individual judgments about what should get excluded when
competition arises. Everyone should agree that widespread causal or rational

---

[4]Scanlon 1998, p. 97.

overdetermination is a problem, but since exclusion arguments often don't do more than suggest their author's preferred way of resolving the competition, they shouldn't be very effective without an agreed-upon procedure for resolving the conflict.[5] There are a number of promising responses to causal exclusion already available, and so seeing Scanlon's objection to the idea that goodness gives reasons as a claim of exclusion is important because it enables us to ask whether similar responses are available in the case of normative reasons. I'll discuss the strategies for avoiding exclusion in the next chapter. Here I simply want to defend the idea that rational redundancy is nothing but rational exclusion.

The fact that it's helpful to think of Scanlon as giving an exclusion argument is no reason to believe it, though. The main reason for thinking this must be Scanlon's argument is that unless it's a claim of exclusion, the conclusion that value never gives reasons doesn't follow. That's one of the things I'll argue in the third section. But first say a bit about what the argument shows, and what it would look like if it were understood as an exclusion argument.

4.1   What does the argument show?

Some philosophers acknowledge a distinction between full-fledged and derivative reasons. Suppose for example that the consideration <I enjoy Alto Cinco's breakfast chorizo> gives me a reason to go there for breakfast. Given that I enjoy it, the fact that <Alto Cinco serves breakfast chorizo> may also give me a reason. But in this context it seems wrong to say that <Alto Cinco

---

[5]c.f. Hooker and Stratton-Lake 2006, who suggest that the redundancy argument is in danger of assuming what its opponents deny, namely that value-making properties are sufficient on their own.

serves breakfast chorizo> gives an additional reason. Whatever force it has is derivative on the force of <I enjoy Alto Cinco's breakfast chorizo>, or so you might think.[6] In the language of the standard conception of normative reasons, we can say that some considerations favor independently and some favor derivatively.

Scanlon's view appears to be that value propositions don't give reasons of any kind, even derivative ones. He says clearly that "being valuable is not a property that provides us with reasons."[7] But it's not obvious that this is the right thing for the buck-passer to say.

Mark Schroeder argues that Scanlon is wrong to think this. Instead, he suggests buck-passers should say that value propositions are excluded from favoring independently, while allowing that they can be derivative reasons.[8] According to the positive buck-passing thesis, value facts are identical to facts about the existence of other reason-giving properties. So considerations of value will be propositions asserting the existence of these other reasons. Schroeder argues that understood in this way, value-considerations do give derivative reasons. He gives following example to illustrate.

Imagine there's a big surprise waiting for Nate in the living room. Nate loves good surprises and hates ruined ones, so we cannot tell him why he *really* ought to go into the living room without taking away his reason. The most we can tell him is that he has a reason to go, which is a value-consideration. Now since acting well means having one's motivating reasons correspond to one's normative reasons, and it's possible for Nate to do well in this scenario by acting on the consideration that <there's a reason to go into the living room>,

---

[6]More examples like this can be found in Schroeder 2011.

[7]Scanlon 1998, pp. 96.

[8]Schroeder 2011, pp. 344-345.

it follows that the value-consideration did give him a derivative normative reason. In this way, value propositions can be seen as place-holders for the value-making considerations that give independent reasons.[9]

This leaves us with two important takeaways. One is that the redundancy argument is probably better thought of as concluding that values don't give independent reasons, rather than the claim that they give no reasons at all. This would still be enough to show that the value-first view is false, since ultimate reasons are independent reasons. The second is that it doesn't follow from the buck-passer's claim that values are higher-order properties of there being other reason-giving properties that values don't give any reasons at all. That is, the negative buck-passing thesis is not a logical consequence of the positive thesis. So buck-passers are not entitled to simply assert the positive thesis and use that as evidence against the value-first view.

Indeed, the negative thesis remains independent from the positive one even when they're re-interpreted by Schroeder's lights. The fact that goodness is the property of there being other independent-reason-giving properties doesn't guarantee on its own that the higher-order goodness property doesn't also give independent reasons. We need the additional premise that such reasons should be thought of as derivative in some way before we can conclude that values don't give independent reasons. We'll return to this point later on.

4.2   Exclusion as a counterfactual

Although it's meant to appeal to more intuitive ideas, talk about redundancy and exclusion is just metaphor. After all, exclusion and redundancy are things that people and other intentional things do, not mind-independent

---

[9]See also Vayrynen 2006, p. 297; and Kagan 1989, p. 60.

metaphysical phenomena. It would be nice to have a more literal way of making such claims. In this section I'll show how we can capture the main thought behind the argument with a counterfactual exclusion claim about what would happen if values were subtracted from the world while their makers remained fixed.

Concerns about exclusion are in the end concerns about what has powers—be they causal or normative—and having a power means making a difference. Counterfactuals are well-suited to express views about difference-making. If x makes a causal difference, then we can say that if x hadn't happened, then things would have turned out differently. And if x makes no causal difference, then we can say that if x hadn't happened, things would have turned out the same. Now, there are well-known problems with using counterfactuals to analyze the concept of causation, but I'm not proposing an analysis. These are just conditionals. The problems for counterfactual analyses arise because some things that are not plausibly *the* cause of some event still meet the counterfactual criteria. It might be true that if I didn't own my car I wouldn't have wrecked it, but that's no reason to think that my owning the car was the definitive cause of the crash. Whether owning the car makes a causal difference may be an open question, but we needn't take a stand on that. The proposal is just that if you have a view about whether something makes a causal difference, that implies a counterfactual.

Normative reasons work in the same way: to count in favor is to make a difference to what we ought to do. And so if x gives reasons, its doing so implies that if x had not been the case and other things were equal, what we ought to do would have been different. Conversely, if x doesn't give reasons, its failing to do so implies that if x had not been the case and other

things were equal, what we ought to do would say the same. If goodness and other values don't give reasons, then this second principle should be true of them. So we can capture the basic idea of the redundancy argument—that values are redundant because value-making properties have done all the work already—with the following *principle of rational exclusion* (PORE):

(PORE): For all values X, value-makers of X, Y, and favored objects Z, if any of the X were subtracted while the set of Y was held fixed, the set of Z would remain unchanged.

More plainly, PORE says that if we subtracted value from the world but held everything else constant that we can, logically speaking, our normative reasons would be the same.

It might seem as though buck-passers can't accept PORE because of their positive thesis that goodness is the property of there being other reason-giving properties. If that's true, one might worry that there simply are no situations in which there could be reasons without there being value as well. Now, whether or not Scanlon and other buck-passers can adopt PORE is a bit beside the point here. This chapter is about turning a loose-talking argument given in metaphors into a straightforwardly intelligible objection to the value-first view of reasons. I'll argue in the next section that PORE is the only way to do this. If the positive buck-passing thesis prevents someone like Scanlon from believing in this principle, so much the better for the value-first view, as its critics won't be able to state their objection clearly. That said, I don't see any reason why buck-passers would need to reject it.

First of all, even if it were logically impossible for there to be reasons

without values, PORE could still be true. Nearly everyone thinks *modus ponens* is logically necessary, but it could still be non-vacuously true that if *modus ponens* were false, all our logic teachers would have been wrong. Even though the nearest worlds where it's false are impossible, it stands to reason that at such worlds the logic would clearly have been wrong. In the same way, it stands to reason that if the nearest worlds where reasons exist without values are impossible, they would be ones where our reasons remain the same, if the redundancy argument is right.

But of course it's very unlikely that anyone thinks of the positive buck-passing thesis as logically necessary. Water and H_2O might not have been identical, just as the descriptions *the morning star* and *the evening star* might have referred to different celestial objects. Identities are generally not considered logically necessary. And even if the positive thesis is analytic rather than logically necessary, it's still consistent with PORE. There's nothing wrong with believing on the one hand that 'bachelor' and 'unmarried man' are synonymous while believing on the other that if they had meant different things, the nearest worlds where all the bachelors fail to shave themselves might be worlds where all the unmarried men still do. So there's no reason why it would be inconsistent to think that x and y are identical and also have a view about what would be the case if they weren't. There are plenty of possible worlds that we can test such a view against.[10]

That aside, there are two related objections that don't hinge on the positive buck-passing thesis. The first concerns the difficulty of evaluating PORE. One might think that since evaluating PORE requires us to make judgments about far-away possible worlds, we cannot say whether it's true or false. The second

---

[10]If you're the kind of philosopher who thinks metaphysical and logical possibility are coextensive, see above.

is that worlds where there are reasons but no values are too far removed from the actual to be normatively relevant. That is, those worlds don't belong in a model of normative reasons for beings like us and so PORE doesn't bear on them—be it true or false. We'll address these more directly in chapter 5, but here it's worth saying a bit about the motivation behind them.

In both cases the primary worry is charity. If PORE is un-evaluable or irrelevant, then it may seem wrong to pin it on Scanlon. Yet charity can become a consideration only when there is more than one option for understanding a view or argument. Because PORE, or something very close to it, is necessary to get the conclusion that values don't give reasons, these objections aren't so much of a reason not to interpret Scanlon this way as they are reasons for thinking that premises of the redundancy argument are false. And we need at least one of the premises to come out false for the value-first view to succeed. That's why we'll postpone discussing them until the next chapter.

4.3   Rational redundancy = PORE

The last section illustrated how we can view the redundancy argument as employing a counterfactual exclusion principle. Now the task is to show that we need to see it that way. There are three points of defense. The first is that PORE is consistent with the goals of people who would cite the redundancy argument as a reason not to accept the value-first view of reasons.

For some, showing that value doesn't give reasons is important because it makes ethics simpler. If the fact that something is good never gives anyone a reason, then we don't have to consider what's good and what's bad when we're

trying to figure out what we ought to do. As Gerald Lang puts it, denying reason-giving powers to value means we can "pour all our philosophical energy into identifying reasons, rather than into identifying values."[11] A.C. Ewing was also someone who thought simplicity is an advantage of denying that values favor. He argued that his view that things are good because it's fitting to have pro-attitudes towards them (and not the other way around) is superior because of the economy of requiring only one basic normative concept, namely fittingness.[12]

If Ewing and Lang are right—and I take no stand on whether they are—the simplicity gained by denying reason-giving powers to values is a point in favor of thinking that PORE is true. When other things are equal, reducing the number of basic normative concepts is a theoretical virtue, and so considerations of theory choice give a defeasible reason to adopt views that can deliver this. And so it should be clear that PORE is compatible with the goal of simplicity. If value propositions make no difference in the sphere of reasons, then ethics can proceed as if there weren't any. Because PORE allows us to be methodological nihilists about value, doing ethics would turn out to be simpler than originally thought, if it's true.

For people like Scanlon who endorse the positive buck-passing thesis, the redundancy argument is an important part of the motivation for that view. Scanlon makes this clear when he says that buck-passing is supported in two different ways: the redundancy argument and the idea that there is no single reason-giving thing that valuable things have in common. About the second motivation, he says:

> ...many different things can be said to be good or to be valuable,

---

[11]Lang 2008, p. 473.
[12]Ewing 1948, pp. 166-169.

and the grounds for these judgments vary widely. There does not seem to be a singe, reason-providing property that is common to all these cases.[13]

The interesting thing about this second motivator is that it's dependent on the first. If values did give reasons, then the property of being valuable would be the reason-giving property that all the instances of valuable things had in common. If it weren't for the redundancy argument, this would be no motivation at all. Thus it's important that PORE be consistent with Scanlon's positive view, and we've already seen that it is. Those who think that value facts are identical to facts about the existence of reasons can still have a view about what would happen if they weren't.

The second point of defense is that PORE would be sufficient for the conclusion the redundancy argument needs, that values don't give reasons. If we take Scanlon's own claim literally, all we get is that:

It is not clear what further work could be done by special reason-providing properties of value, and even less clear how these properties could provide reasons.[14]

If it's simply unclear how values could give reasons, that's not really an objection to the value-first view. It might be that we can be sure that values do give reasons, but we're still be unsure *how* they do, given the rational competition problem. This is a small point but it's important. Scanlon and others who champion the redundancy argument have presumably had something stronger in mind all along, but it's important to make that explicit. PORE does that well.

---

[13]Scanlon 1998, pp. 97-98.

[14]Scanlon 1998, pp. 97.

This brings us to the third point of defense, which is that none of the obvious alternatives to PORE are able to deliver the conclusion that values don't give independent reasons. As I see it, the alternative features of value that might do the work of showing why it doesn't give reasons are these.

(a) ontological inheritance

(b) causal inheritance

(c) being higher-order

(d) explanatory inheritance

(e) giving instrumental reasons

We'll look at each of these in turn, and I'll argue that none of these can work as the thought behind the redundancy argument. Some are just not true of value, and the rest fail to deliver the conclusion we need.

4.3.1 (a) Ontological inheritance

Ontological inheritance is a common notion in philosophy. Plato held the view that things inherit their reality from the forms they instantiate. Others claim that higher-level scientific properties, like the properties of geology, are realized by their lower-level counterparts in chemistry and physics. And many think that when some clay composes a statue, the statue depends for its existence on the clay, to name just a few examples. As we've seen, goodness and other values are realized in this way by lower-level, non-normative value-making properties. If we understood the redundancy argument as appealing to ontological inheritance, then it would go as follows. Reasons given by value would be derivative at best because value properties could not exist without the

non-evaluative properties they depend on, and so the reasons they give could not exist without those other properties. Reasons given by value-making facts would be independent because they aren't ontologically dependent in this way.

Yet this explanation seems to get things backwards, at least by comparison with other cases of ontological inheritance. As a rule, x's inheriting its existence from y doesn't interfere with x's ability to give meaningful reasons. Consider the standard example of Lumpl the clay, and Goliath, the statue composed by Lumpl. Goliath depends for its existence on Lumpl in the relevant way, but it is far from clear that reasons given by Goliath should be derivative or less meaningful just because of that dependence. If Goliath is a prized statue, it will provide reasons for its preservation that are at least as meaningful and genuine—if not more so—than the reasons Lumpl gives. If any reasons are derivative in this case, they are the reasons Lumpl provides. Or, consider mental states like pain. The state of a person's being in pain plausibly depends for its existence on lower-level chemical states of the brain. But this is no reason to think that the reasons a person's pain gives us to alleviate it or take account of it are derivative or less meaningful. Again, if there are derivative reasons here, they are the reasons given by the lower-level brain states.

Now compare the situation if a principle like PORE were true of statues and mental states. I don't think any such principle is true, of course, but suppose for the sake of argument that it were. Then subtracting statues or pain from the world, while keeping Lumpl and the lower-level states, would make no difference to what we ought to do. That alone is sufficient to show us that any reasons given by statues or pain are dependent on the lower-level states. So, an exclusion principle would give a much better explanation of why

statues and pain are normatively inert (if that were a fact to be explained) than an appeal to their ontological dependence. This suggests that PORE gives a much better defense of the idea that reasons given by value are derivative than an appeal to ontological dependence. Values do inherit their existence from their makers, but that alone is insufficient to show that they favor derivatively or not at all.

Some may object that the statue and pain cases are importantly disanalogous to values. Perhaps not only do values inherit their existence, maybe the the derivative reasons they provide do too. These reasons would depend for their existence on the independent reasons given by non-value facts on which the values depend. According to our new explanation, it would be this inheritance relation between reasons, rather than the one between values and their makers, that's responsible for exclusion. This is in contrast to the statue and pain cases.

But swapping out the things related by ontological inheritance doesn't help much on its own. On the new view, the argument must be that the reasons given by values are derivative precisely because they inherit their existence. Now compare that rationale with our statue and pain cases. It's not obvious that a statue is less fully an object just because it's ontologically dependent, or that pain is less of a mental state. Likewise, I don't think an ontologically dependent reason would be any less of a reason. Of course, people have given arguments for deflationism or nihilism about composite objects and eliminativism about mental states on the basis of their relationships with their lower-level realizers.[15] And the fact that those views need sustained arguments with additional premises shows that it's not obvious. Those who adopt this

---

[15]See for example van Inwagen 1990, and Churchland 1981, respectively.

revised view will have to claim that the redundancy argument is doing the same thing as those arguments, and that seems implausible given what Scanlon actually says.

4.3.2 (b) Causal inheritance

Alternatively, one could resist PORE as the center of the redundancy argument by claiming that values inherit their causal powers. Like ontological inheritance, causal inheritance is often put to work in arguments for deflationism or nihilism about some domain. Trenton Merricks argues that composite things would have to inherit their causal powers from the simples that make them up in his baseball argument for compositional nihilism about material objects.[16] And Jaegwon Kim has argued that if mental states inherit their causal powers from their neurological realizers, that casts doubt on the idea that mental events do not reduce to those realizers.[17] If values inherit their causal powers from non-value facts, then this could be the idea behind rational redundancy. This would be a causal exclusion claim about reasons—because all the causal work is done by non-value facts, there's something suspicious or derivative about reasons coming from values.

Similar arguments to the ones above will dispatch the idea that something's inheriting its causal powers shows us that any reasons it gives would be less genuine or derivative. Consider the case of pain again, and suppose that conscious mental states like pain inherit their causal powers from their lower-level neurological realizers. As we've seen, it won't follow on that basis that the state of pain gives less genuine reasons.[18] Compare again what an

---

[16]Merricks 2001.

[17]Kim 1991.

[18]See for instance Fodor, 1989.

exclusion principle would show. If someone's being in pain makes no difference to what we ought to do, then any reasons given by pain do appear suspect. And as before, if the causal exclusion claim is instead meant to say that there's something ontologically—rather than normatively—suspicious about reasons provided by goodness, the interpretation of the redundancy argument we're left with seems at odds with what advocates of the redundancy argument appear to mean. Plausibly, many ordinary things inherit their causal powers, but it takes much more than that fact to cast doubt on them. PORE can license the inference to the conclusion that values don't give independent reasons without the additional cost.

4.3.3 (c) Being higher-order

Some might resist the equation of PORE with redundancy by appealing to simplicity. Perhaps the redundancy argument just needs to be understood as a consequence of the positive buck-passing thesis: value facts are higher-order facts about the existence of other reason-giving properties, and higher-order facts contribute nothing over and above the lower-order facts they depend on. This reasoning seems natural enough. The higher-order fact of there being causally efficacious properties would have nothing causal to add over the powers of its first-order realizers. And the fact of having mass does wouldn't contribute to a thing's mass, and so on.

This line of reasoning might be tempting, but it's not a good candidate for understanding Scanlon's argument. If buck-passers lean on the positive thesis to defend the redundancy argument, they will be undercutting the only independent motivation they have for it, as we saw above. The claim that value facts are higher-order is something buck-passers need to argue for, not

evidence for the claim of redundancy.

Of course, critics of the value-first view need not find themselves in Scanlon's dialectical position. One could have independent reasons for thinking value-facts are higher-order in this way. Still, this won't do. Schroeder's surprise party example showed that higher-order facts do give reasons, and we would need some further premise to get to the conclusion that those reasons are derivative in some way. The cases of higher-order facts about causation and mass could be thought to provide evidence for the idea that higher-order facts are always derivative, but there's an important difference between those cases and normative reasons. The relations of causing and adding mass are extensional, while the relation of counting in favor is intensional. So absent a reason to think this difference is irrelevant, the critics are still missing a premise.

4.3.4   (d) Explanatory inheritance

Advocates of rational redundancy could instead hope to shore up the argument with the idea that values inherit their explanatory power from value-makers, and that explanations from value alone are uninformative. This is explanatory inheritance. Plausibly, specific instances of general laws inherit their explanatory power in this way. My explanation that a particular bird has blue feathers on its wing because it's a drake Mallard gets its explanatory force from the general rule that all male Mallards have such plumage. Without the rule, my explanation fails to inform. If this is true of value then any explanation from value would have to be shorthand for the explanation the relevant non-value facts provide. So on this view, to explain that I ought to choose a certain item from a restaurant menu because it's good would be to

give the very same explanation as that I ought to choose it because it tastes a lot like something I enjoyed as a child, and because it's low-fat, and so on.

This may look like a promising account of the redundancy argument, since to give a normative reason is to give a kind of explanation. But there are a couple of reasons why explanatory inheritance won't work. The first is the argument from chapter 3. If the Guise of the Valuable is right, non-value considerations are never the ultimate explanation of our motivations. And if Motivational Ought Implies Can is true, the very possibility of practical wisdom rests on the fact that there can never be best and most ultimate reasons that aren't possible ultimate reasons. The best reasons, and therefore the best explanations of why should do or think something, are value considerations.

Relatedly, open question considerations cast doubt on the idea the values inherit their explanatory power. Appealing to value can add something to explanations beyond what appeals to non-value facts can offer. Depending on the context and background beliefs of my dinner companions, my explanation that I ought to choose the meal I did because of its similarity to food I enjoyed as a child, because of its fat content, and so on, may fail to be informative. In spite of the information I've given, it can be an open question why I should choose that dish. Conversely, if I explain that I ought to choose it because it's good, the question no longer seems open. Scanlon himself cites this phenomenon as the thing that led him to consider the buck-passing theory.[19] This seems like a good reason to deny the explanatory inheritance claim. If critics of the value-first view endorse PORE instead, they can underwrite their argument without needing to confront these.

---

[19]Scanlon, 1998, p. 96.

4.3.5 (e) Giving instrumental reasons

Our final option lies with instrumental reasons. One could try to fill out the redundancy argument with the idea that the reasons values give are merely instrumental, as opposed to the basic or final reasons given by non-value facts. If instrumental reasons are derivative or superfluous, the ends given by non-value facts would give all the independent reasons there are. Thus it would be no wonder that values don't give independent or ultimate reasons, and that ethics can proceed as if they didn't exist.

Once again, this is a tempting line of argument. But this way offers little improvement over the others. The problem is that instrumental reasons do not appear to be derivative in the relevant sense. One way of explaining instrumental reasons is that they're reasons only to the extent that they work in service of a further end. But that's importantly different from the idea that ends or their qualities give a complete explanation of instrumental reasons, which is what this version of the redundancy argument would require.

For example, suppose one of my ends is to satisfy my desire to climb the peak of some difficult and remote mountain, although my desire allows for some latitude in how that's accomplished. And suppose that there are three routes to the summit, offering varying prospects for success. I know that if I take Route A, the probability of my reaching the summit successfully is 7/8, if I take Route B, the probability drops to 1/2, and there is a 1/5 probability if I go by Route C. Assume I have no other ends that bear on which route to take. In this case I have instrumental reason to choose Route A, but that reason cannot be explained only by my desire to reach the summit. To explain it, we need to appeal to the further fact that Route A is the most likely to lead me to my goal, and so the instrumental reason I have for choosing that way

is non-redundant. Similar arguments would show the same for many other instrumental reasons. So, the idea that the reasons given by values would be instrumental to the ends given by non-value facts is no argument for the claim that they are redundant.

To avoid this conclusion, advocates of redundancy would need to show that the assumption that there are no other ends that bear on the decision among routes is unallowable, that there are *always* other ends that work together to single out one option. One might think that people typically adopt the end of pursuing their other ends in the way that's most expedient, to the extent that it's compatible with the other reasons they have. And perhaps second-order ends like are what make instrumental reasons redundant.

This strategy is not very plausible, though. The second-order end of expediency is not a good candidate to be an end in the first place, because it can never be satisfied except in cases where a person has attained all their goals and never adopts any more. Instead, we should view the norm of expediency in practical reasoning as expressing the fact that some reasons are instrumental rather than basic. In spite of the fact that we would not have any instrumental reasons were it not for the ends that we have, norms of practical reasoning like expediency show that there are some practical problems that ends alone do not settle, or account for. And that is compatible with allowing that instrumental reasons are non-redundant.

So to conclude, PORE appears to be the best option the friend of the redundancy argument has for making the argument a straightforward objection to the value-first view of reasons. And if that's right, we now have a basic strategy for mounting a response. If there are good reasons all things

considered to believe PORE, then the value-first view is probably wrong. If not, then it might have some promise after all. We'll pursue this line of defense in the next chapter.

Chapter 5

Against Rational Exclusion

Think of a funny joke. Now imagine that all the things that contribute to the joke's comedic value—like context, delivery, and psychology—were held fixed, but the joke no longer had the property of being funny. That's an odd situation, but it's coherently imaginable. The kind of question we're going to consider in this chapter is whether the reasons for laughing at the joke change when we hold the comedy-makers fixed and take the value away. If the answer is no, then *being funny* or *being a good joke* are rationally excluded. That is, those values don't give reasons. We'll consider this example, along with other cases involving of normative reasons, to help decide whether value is excluded across the board.

In chapter 4, I argued that Scanlon's redundancy argument against the idea that values give the best and most ultimate reasons (the value-first view), is at bottom an exclusion objection. The goal for this chapter will be to defend against that objection by arguing against the principle of rational exclusion (PORE). If this succeeds, then the most significant obstacle for the value-first view will have been removed.

Think of the redundancy argument as having two premises:

(a) PORE, and

(b) if PORE, then values don't give reasons.

I'll argue that both of these premises are questionable, though we'll focus more on (a) than (b). First I'll suggest that even if PORE were true, it might make no difference to the reasons we actually have. This would mean that premise (b) is false. Then I'll show how PORE appears to create a regress in which all our normative reasons would disappear. The same basic thought in PORE that excludes value will also exclude most familiar things from giving normative reasons. And that means that most if not all of the considerations that give reasons are considerations we cannot understand, and thus not reasons at all because the very possibility of practical wisdom rests with the idea that we can act for the right reasons. This would make premise (a) false. Either way, the redundancy argument fails.

Before we look at those arguments, let's review how we got here. The value-first view is a solution to the rational competition problem, which arises because common sense wants to allow that both values and their makers give reasons for the same things. That this is the right solution to the problem is implied by the versions of Ought Implies Can and the Guise of the Good we saw in chapter 3. The value-first view says that every ultimate normative reason is a consideration of value, where values are are the typical outputs of standards of evaluation for actions and attitudes. So our best and most fundamental reasons are always considerations of the form $<x$ is F-ly good$>$ or $<x$ is F-ly bad$>$, which is precisely what Scanlon's redundancy argument

denies. That's why the argument has to be met for the value-first view to come out looking defensible.

So much for introduction. In the next section we'll consider whether PORE can be reliably evaluated and whether it makes any difference to our actual reasons. Then we'll tackle PORE itself, and lay out the regress argument against it. If what I argue is right, concerns about exclusion should no longer threaten to undermine the value-first view of normative reasons.

5.1   The relevance of PORE

Roughly, PORE says that if we subtracted value from the world while holding everything else fixed that we logically can, our normative reasons would remain exactly the same. We can capture that idea more precisely by saying:

(**PORE**): For all values X, value-makers of X, Y, and favored objects Z, if any of the X were subtracted while the set of Y was held fixed, the set of Z would remain unchanged.

One immediate difficulty is that this claim is pretty abstract. It can be difficult to think clearly about what things would be like if the antecedent of PORE were true. That's where the joke case from above comes in handy. That thought experiment is a test for a specific instance of PORE. We took a joke, took the comedic value away, held fixed everything else, and then asked what our reasons would be. If our reasons with respect to the joke would be exactly the same, that means that comedic value—that is to say, the property of being funny—is excluded from giving reasons. And if our reasons would

change, that means that PORE is false when it comes to reasons of comedy. If this kind of value makes no difference to the existence of reasons, then it certainly doesn't give them. The same goes for all the other kinds of value. If moral values, epistemic values, prudential values, aesthetic values, and the rest make no difference to the distribution of reasons, then they don't give them. But if they do make a difference, then the claim of rational exclusion can't be the reason that value considerations don't count in favor of things, because the claim is false.

To be clear, this way of combating the exclusion objection does not offer a way re-affirm that values do give reasons, as a sort of back-up argument to the one from chapter 3. Evaluating PORE won't be able to show this. If subtracting value while holding everything else fixed did change the reasons we have, all that shows is that values make a difference to what reasons there are. And while anything that gives reasons would make a difference in this way, simply making a difference is not sufficient for reason-giving. That's because there are other ways to make a difference to what reasons there are.

For an obvious example, consider the existence of the universe and the existence of rational beings like us. These make a difference to nearly all of the reasons there are—if there were no universe or no rational beings like us, then the reasons we have not to torture people would disappear. But that fact is hardly evidence for thinking that the existence of those things give reasons against torture. These are rational background conditions, so to speak, rather than actual reason-givers. This point mirrors a common objection to difference-making theories of causation. Analyzing causes in terms of difference-making dramatically over-generates causes, so the objection goes. The fact that I have bones makes a difference to the breaking of my arm,

but it's not the cause of the break. And the fact that my house is not currently on fire makes a difference to whether I'm writing my dissertation at this moment, but again is not the cause of my writing.[1] Given the analogy between normative reasons and causes, this is likely not surprising.

So, thought experiments like the joke case can help us to test PORE, but they can't weigh in on the value-first view more generally. Here are some other examples.

- *Aesthetic Reasons*

  Imagine a good work of art and the reasons it generates. Now imagine subtracting whatever aesthetic value it has (its beauty, perhaps) while holding all of its good-making qualities fixed. Are our aesthetic reasons the same as they were before the subtraction?

- *Moral Reasons*

  Take something that's morally wrong (or morally bad) and consider the reasons we have for preventing it. Then suppose that it's no longer wrong, but its wrong-making features remain, and consider whether our moral reasons for avoiding the thing will change.

- *Epistemic Reasons*

  Think of beliefs and the reasons that justify them. Then ask whether the epistemic reasons for believing or disbelieving them would change if the justification were removed while the properties that made for justification remained fixed.

- *Prudential Reasons*

  Consider something prudentially good, maybe an efficient means to your

---

[1]See similar points in Menzies 2004.

ends. Will the reasons you have for pursuing this thing change if the prudential goodness is taken away but the efficiency remains?

There are other kinds of normative reasons and value to run this thought experiment on: legal reasons, reasons of etiquette, and probably others besides. For PORE to be true, it has to be that all of these cases are ones in which the reasons would remain unchanged. The argument from chapter 3 already shows that we should judge the reasons to be different in all of the cases PORE covers, but defeating the exclusion objection requires only one of the thought experiments to go that way.

Bringing the test down from the level of generalized value and reasons to more specific kinds may help make PORE easier to think about, as the more determinate tests are closer to our everyday thinking about values. Of course, some might find generalized PORE easy to decide its own. I'll continue to speak about the more specific versions of the test to make the discussion a bit more gripping, but obviously everything said about those will go for PORE itself. They're just ways of testing PORE.

Now, even if testing PORE by way of thought experiments on more determinate kinds of value can make things easier, it still may not be feasible to decide. Consider the epistemic and aesthetic versions of the test for instance. I have a justified belief that I'm sitting in front of a computer right now. And (let's suppose) the Miles Davis that I'm listening to while I type is a great work of art. On the one hand there is a pull towards saying that if everything were held fixed except for the facts that my belief is justified and that *Kind of Blue* is a good piece of music, my reasons for believing it and appreciating the music would stay the same. Why wouldn't they? I still have all the same evidence, and the music sounds exactly the same. On the other hand, there's

the upshot of chapter 3: if it isn't valuable in some way, you shouldn't do it. And that suggests that if the belief is no longer epistemically good and the music is no longer aesthetically good, then the situation with my reasons might well have changed. That doesn't mean on its own that values aren't excluded, but it's enough to give pause when we think about these cases. The redundancy argument challenges us to make a seemingly difficult decision here. That's why the question of rational exclusion is interesting and philosophically significant.

There's a possibility that PORE is not just hard to make up one's mind about. It might be impossible to decide, un-evaluable in principle. As we saw in chapter 4, interpreting the redundancy argument along these lines might seem uncharitable, effectively suggesting that the argument amounts to near-nonsense. But as I argued there, it should be clear that PORE really is evaluable. It's not logically impossible to hold value-makers fixed while subtracting value. Since the antecedents in all the counterfactuals we're considering are logically possible, there's no reason to worry that they aren't evaluable, even if they are hard to decide on.

Still, the difficulty of assessing PORE and its more specific instances is some evidence for thinking it's too far-fetched to make any difference to actual-world reasons. It may be entirely irrelevant what gives reasons, even if it's evaluable in principle. In that case, the second premise of the redundancy argument would be false.

Consider the counterfactual *if modus ponens were invalid, all the logic books would be false.* This strikes me as both non-vacuously true and easy to decide, but its truth tells us nothing about what kinds of results are derivable from

modus ponens.[2] Maybe PORE is like that. It might be that what's true or false at distant worlds that are just like ours but for the existence of value simply make no difference to what actually favors. Nearly all of the situations we consider in ordinary normative thinking are ones where supervenience holds. Think of trolley cases, or cases about what our reasons would be like if x had happened rather than y. But the worlds we need to evaluate PORE are ones where supervenience fails, since in all of them we have a difference in value without a difference in value-makers. It doesn't seem obviously wrong to think that because supervenience is a foundational commitment of normative theorizing, any worlds that violate it are too far removed from the actual world to matter. And whatever difficulty we might feel in evaluating the relevant counterfactuals would be nicely explained by the irrelevance of such worlds: we're so accustomed to reasoning with supervenience that its failure leaves us confused.

The concern that some thought experiments are normatively irrelevant is predictable, and *prima facie* compelling. An example from Parfit's *On What Matters* can illustrate this. Parfit gives an argument against subjectivist theories of normative reasons. According to this kind of subjectivism, reasons are given just by facts about our current desires. This is far too quick a statement of the view, but the details don't matter much here.[3] Parfit asks us to imagine people who have no desire to avoid agony. He claims that such people certainly do have reasons to avoid agony, and subjectivism implies that they do not. So subjectivism must be false.[4]

---

[2]Adherents of the orthodox semantics for counterfactals will balk here, as counterpossibles like this one are supposed to come out as vacuously true. For a good argument that the orthodoxy gets this wrong, see Nolan 1997.

[3]See Parfit 2011, part I, chapter 3 for more discussion.

[4]Parfit 2011, p. 76.

One response to this argument is that Parfit must be wrong because subjectivism is not a theory about people who are so radically different from us. The idea here seems to be that if our theories of normative reasons aren't meant to apply to people who don't care about something so fundamentally bad for us as agony, then Parfit's argument has committed the fallacy of irrelevance. The reasons such people would have should make no difference to the reasons we actually have, or so the objection goes. Parfit seems to take this sort of response seriously, since the first objection he considers is that the case is "purely imaginary," and that "every actual person ... wants to avoid agony."[5] Whether this objection to Parfit succeeds is indifferent for present purposes. But the example shows that concerns about irrelevance ought to be taken seriously.

So, the if-then premise of the redundancy argument is weak because the truth of PORE may not tell us anything about what actually gives reasons. But there are also good reasons for thinking that if PORE were relevant, it would have to be false. We turn now to the principle itself.

5.2 Does favoring drain away?

In this section we'll unravel PORE by pulling on the very thread it's made of. On the assumption that it's both reasonably evaluable and relevant to what actually gives reasons, the following problem emerges. If the intuitive motivation for excluding values—that the value-makers would have been sufficient on their own—is correct, a regress looms in which favoring powers are pushed down to the lowest level of ontology. That's a bad thing, as I'll explain.

---

[5]Parfit 2011, p. 76.

In our post-Moore philosophical culture, it's easy to get into the habit of thinking that there are just two levels of description needed to do ethics: there's the level of goodness and the level of good-makers. And when those are the only levels that matter, excluding value in favor of its makers can look like the right choice. Value-makers are normatively efficacious by their very nature, since they bring about value. So when the makers compete with value to give reasons, it can seem like good sense to resolve the competition in favor of the makers—it's a conceptual truth that the makers have normative powers, but it's less clear that values themselves have them.

This story about exclusion might seem like a misrepresentation of Scanlon's rationale. He argues, after all, that the best explanation of the openness in Moore's open question argument is that although the claim that x is F might leave it open whether we should draw some conclusion about our practical reasons, the claim that x is good does not leave it open—if it's good, we ought to have pro-attitudes towards it.[6] So Scanlon does think that values can help us find the normative reasons that we have. But being a guide to the existence of reasons is not the same as having normative powers. Scanlon thinks that values are indicative of the ability to favor because they are generated by favoring. That's not the same thing as having genuine normative powers.

The beginning of the problem for PORE is that exclusion cannot stop at the level of commonsense value-makers like *being pleasant* or *shedding light on the causes of cancer* without doing so arbitrarily. It wouldn't be arbitrary to stop at the level below value if there were only the two levels. But there's a lot going on below the level of *being pleasant*. For each level of description—psychological, neurological, chemical, physical, etc.—there is

---

[6]Scanlon 1998, p. 96.

a corresponding group of properties on which commonsense value-making properties supervene, just as value supervenes on the makers. Consider the reasons given by these commonsense value-makers. To the extent that qualities like *being pleasant*, *being a belief caused in the appropriate way*, *maximizing happiness*, or *looking like Guernica* give reasons, it seems clear that they would do so in virtue of lower-level qualities. A pleasant resort is desirable because of its basic physical features like weather and location, an appropriately caused belief is desirable because there is a truth-conducive physical connection between the believer and the object of belief, maximizing happiness is desirable because of the neurological condition that happy people are in, and having the appearance of *Guernica* is desirable because of photons entering people's eyes and the kind of neurological response they tend to generate. Thus the higher-level value-makers supervene on lower-level makers, and there is reason to suspect that these lower-level makers are sufficient on their own to generate all the reasons we have.

As we saw in chapter 4, supervenience of powers and a presumption of sufficiency are all we need to generate an exclusion argument. So if the exclusion thought behind PORE works the way it's supposed to, it should also work for the conclusion that higher-level value-makers are excluded in favor of the lower-level makers. Exclusion operates on the idea that when two different levels related by supervenience are sufficient to produce the same outcome, the competition between them should be resolved in favor of the lower-level thing. Again, when we're only thinking of two levels, we've got no choice but to stop after one iteration of the exclusion argument. But when there are many levels, it's arbitrary to stop after just one level—you have to ride the elevator all the way to the bottom, so to speak. In other words, either

you grant powers to the top level or you give them to the lowest. There is no tenable middle ground.

The idea that exclusion must run all the way down to the bottom level comes from Ned Block, who invoked it to combat Kim's causal exclusion argument against non-reductive physicalism about the mind.[7] Block argues that if the mental is excluded from having causal powers because its neurological supervenience base is already sufficient, then the neurological is excluded in favor of its chemical supervenience base, and so on down to the lowest level of physics. Block goes on to suggest that there may be no bottom level of physics. It might be turtles all the way down.[8] And if that's the case, then causal powers drain away in a regress that never stops, which means that nothing really has causal powers. That's a bad result for fans of the causal exclusion argument.

The present argument against PORE adapts Block's rationale to normative reasons, but it differs from his argument in one important way. Block's conclusion that causal powers drain away into nothing works only to the extent that it's likely that physics has no bottom level. It's possible to run the argument that favoring powers drain away with this same premise, arguing that if there's no bottom level of physics then nothing can give reasons because favoring powers will be stuck in the same regress as causal powers. But the "no lowest level" premise is making a substantial assumption, which is subject to empirical confirmation. If it turns out that physics really does have a bottom level, then Block's argument is no good. This is where our argument against PORE departs from Block's: it doesn't need any claims about whether or not there is a lowest level of physics. Because favoring can't happen below a

---

[7]Block 2003.

[8]Because of a promise I made to Matt Koehler, let's stipulate that these are ninja turtles.

certain threshold, and because the regress will push the eligible considerations below it, there wouldn't be any normative reasons at all. Hence, if PORE is true, favoring might just drain away.

Before we detail how that happens, there are a couple of issues to mention. The first is a reminder that thinking of higher-level properties as identical with constellations of lower-level properties is perfectly fine for our purposes, but it won't help the PORE out of the regress problem. That's because PORE is concerned with considerations, which are facts or true propositions, rather than properties. As we saw in chapter 4, even if the properties *being Lumpl* and *being Goliath* are identical, that doesn't mean that considerations about Lumpl and Goliath give reasons equally. That some statue is Goliath could plausibly be taken to give reasons for displaying and preserving it, which the fact that some bit of clay is Lumpl would not plausibly give. Generally speaking, it seems that considerations count in favor *de dicto*, rather than *de re.*

For example, if I'm Lex Luthor and I want to kill Superman, the fact that Clark Kent works at the Daily Planet doesn't seem to favor any course of action for me unless I also know that Clark is Superman. So even if higher-level value-makers are identical in some way to lower-level makers, that shouldn't make any difference to whether the higher-level makers can be excluded. Reason-giving powers can't be inherited just by dint of identity. Just as people worried about epiphenomenalism were concerned about whether our actions are caused in virtue of the mental qualities of our brain states or whether they're caused in virtue of their physical qualities—even if the relevant physical and mental states are identical—so too does it seem that it wouldn't matter if value facts and value-maker facts were really the same. If PORE is true, then facts favor in virtue of their value-making qualities, not their value

qualities.[9]

The other thing to mention is that it's unlikely that Scanlon, or anyone else who accepts the redundancy argument, meant for the exclusion thought to be iterated into a regress. Indeed, since the publication of the original redundancy argument Scanlon has said that lower-level evaluative properties like wrongness do in fact give reasons.[10] Now, it's unclear what this means. Scanlon is of course entitled to the positive view that lower-level evaluative properties do give reasons, but it's far from clear that he can hold this view while embracing rational exclusion at the same time, as we saw in chapter 2. So as Roger Crisp points out, this may be a change to Scanlon's original view, or it may not be.[11] If Scanlon is giving up the redundancy argument, then a lot of the wind goes out of its sail. But if not, then stopping the buck anywhere between the top level and the bottom level is unacceptably arbitrary in the absence of a good reason to do so. This would be the case regardless of whether the buck stopped at lower-level evaluative properties, high-level natural properties, or mid-level natural properties. I mention this only to make it clear that drainage of favoring powers isn't a feature of PORE, it's a bug.

With those things in mind, it's time to put the argument into sharper focus. The basic claim is that if PORE is true and relevant to what actually gives reasons, then the line of reasoning that excludes values must also exclude everything down lowest, most determinate level of physical description. Here are a few examples of how that happens, utilizing the specific applications of PORE that I described above.

---

[9]For a quick history of how this came to be the exclusion thought, see Fodor 1989.

[10]Scanlon 2002, p. 513.

[11]Crisp 2005, p. 83 n. 6.

- *Aesthetic drainage*

  Suppose we took considerations of beauty to give reasons. Aesthetic PORE says that if we subtracted aesthetic values from the world but fixed everything else we logically can, there would be no change in our aesthetic reasons. So at best, higher-level aesthetic good-makers, like *looking like Guernica* are what give the reasons. But those higher-level makers have mid-level makers, like *being this arrangement of paint on canvas.* Once exclusion is allowed, there's nothing to stop us applying a version of PORE for higher-level aesthetic good-makers such that if we subtracted everything down to and including that level while holding fixed everything else, our aesthetic reasons would stay the same. The same will go for arrangements of paint. If the higher-level makers would have been sufficient, then so should the middle and lower-level ones. We can continue to apply this procedure until we've excluded everything but constellations of the most basic properties of paintings there are.

- *Moral drainage*

  Moral PORE prevents moral values like wrongness from giving reasons because there are moral value-makers, like non-compliance with the Categorical Imperative (CI), that are presumed to be sufficient to generate all the moral reasons we have. But of course, non-compliance with the CI has its own makers, something along the lines of being in a particular type of neurological state. Taking away everything down to the neurological level while holding the rest constant would presumably leave us with the same reasons too. So compliance can't be what favors either. Again, we can repeat this inference for successively lower levels until we reach the bottom.

- *Epistemic drainage*

  Reasons for and against believing things have to come from somewhere, but if Epistemic PORE is right, then they won't come from an epistemic value like justification. At best they'll have to come from a maker of justification, like *being causally connected with the object of belief in the appropriate way.* That connection will itself have makers at lower levels in increasingly determinate relations among properties and objects. So if PORE works against justification, similar thoughts will work against causal connections and the less determinate relations that make for those connections.

- *Prudential drainage*

  Prudential goodness and badness are excluded by PORE in favor of lower-level properties like *being the more efficient option to one's ends*, since they are thought as sufficient on their own. And as in the other cases, efficiency is realized by more complex lower level properties. In this case, those realizers will be a hyper-determinate combination of logical and causal relations. Once exclusion gets started, the only things left to give prudential reasons are these maximally basic properties.

- *Comedic drainage*

  Applied to comedy, PORE tells us that the reasons we have to laugh at jokes are already accounted for by properties that are of a lower level than *being funny* or *falling flat.* You might think that something like *being a clever pun* would suffice. But there's nothing to stop us presuming that properties that the sub-vening properties that make for cleverness would also be sufficient, thus excluding even cleverness, etc., until we reach basic facts about how the brain interprets things that are communicated

to us.

Hopefully these examples are enough to illustrate how good-makers of every kind are vulnerable to drainage of favoring powers, by virtue of the fact that they all have their own realizers. Since the mechanism of exclusion is a presumption that realizers would be sufficient on their own, this will extend all the way down until we reach a level where there are no more realizers.

The only way out of the regress is to find a point where certain upper-level good-makers would be sufficient on their own, but the realization base at the next level down would not be. If that were the case, then favoring powers wouldn't drain all the way down to the lowest level because the train of exclusion would stop once it reaches a point where lower levels are no longer sufficient.

As we saw above, Scanlon explains Moore's open question argument as follows. While it can be an open question whether the possession of certain value-making properties ensures that there are reasons in the vicinity, the possession of value itself never leaves that question open. It can be tempting to think something like this will do the trick to stop rational drainage: if some group of higher-level value makers settled the question of whether there are reasons towards the things that have them, but lower-level makers left the question open, that offers a natural breaking point to stop the iteration of exclusion. The thought would be that the higher-level properties are obviously sufficient to generate the reasons we have, but the lower-level ones are not. Exclusion arguments break down once the supervenience base is no longer sufficient on its own to generate the outcome in question.

The trouble with appealing to open question considerations to stop drainage in this way is that leaving the question of whether there are reasons

in the vicinity of some property open doesn't mean that the property is insufficient on its own to generate those reasons. Which questions are open and which one's aren't is a matter of conceptual necessity, but the reason some consideration gives need not be conceptually necessary. Scanlon's own example can illustrate this. The property of *being pleasant* leaves the question open according to Scanlon, but it's still sufficient to generate reasons in his view.[12] Scanlon's later claim that pleasantness is really a lower-level evaluative property complicates things, but the general point stands. Even if *being pleasant* also settled the question conceptually, that should not stop the realizers of pleasantness from being sufficient to give reasons with something less than conceptual strength. So an open question argument won't do to stop favoring powers from draining all the way down.

Another way to stop the regress might be to use the threshold of psychologically possible motivation, of the kind we saw in chapter 3. If reasons were given just by properties at the lowest level we can appreciate, then they wouldn't be in danger of draining away into nothing. There are a couple of things to say about this strategy. The first is that we need a substantive proposal about where this border is in the spectrum of levels. Since claims of psychological possibility can't be made universally, there may be no clear line of division between what's possible and what isn't. Without a firm line of demarcation it could be possible to run exclusion arguments on pairs of properties at adjacent levels, and push the regress through to the other side, as in a *sorites* series. And even if the regress can't be pushed through in this way, the potential vagueness of the border would leave us with no clear sense of which properties are the genuine reason-givers, and our ability to act on the

[12]Scanlon 1998, p. 96.

right reasons would still be hampered. So an objectionable kind of rational drainage would nevertheless occur.

The second response to this move is that psychological possibility doesn't seem to capture the intuitive thought behind PORE. Exclusion gets going on the idea that as long as some set of properties is materially sufficient to give the result we want, they get the powers in question. No appeal to psychological possibility is needed, and Scanlon certainly doesn't make any. One of the chief concerns about the redundancy argument is that it runs together the metaphysical explanation of how reasons occur with the normative story of why those reasons are reasons. It's true that the lowest level of ontology is sufficient to produce all our reasons, and insofar as that's the ticket to giving reasons, the regress should continue.

On its own, the idea that all the favoring is done at the lowest level of description might look attractive to those who appreciate hard-headed scientific realism. This is a sort of normative analogue to nihilism about composite material objects. Where nihilism claims that only the most basic things do the existing, this is the claim that only the most basic considerations do the reason-giving. But this is just the first half of the problem for PORE. The criticism of PORE is not just that reason-giving powers drain down, it's that they drain away. That is, they disappear. And that should be much less palatable to anyone who's sympathetic with normative realism.

If favoring powers drain down to the lowest level, then the considerations that actually favor in the end will be hyper-determinate and maximally complex. In many cases they will be considerations about arrangements of sub-atomic particles, quantum field states, and the laws that govern their interaction. More specifically, they will be propositions quantifying over vast

numbers of things and involving more than a few relations. Think of a proposition like <there is an x, a y, and a z, such that Fx, Gy, and ¬Gz, and Rxz and ¬Ryz, or there is an x_1 such that Syx_1>. That's already complicated enough for most of us, even when the predicate symbols mean something ordinary and accessible. Now imagine a proposition like this where the predicate meanings are supplied by the most fundamental particle physics we have, and which has millions more argument places. That might sound a bit overdramatic, but that's quite literally the sort of proposition that we should expect to find at the bottom level.

Most of us, and more likely all of us, lack the brain power to fully consider such a proposition, even if we understand the relevant physics. These considerations are simply unappreciable for beings like us. And that means it's not psychologically possible for us to be motivated by them in the sense I introduced back in chapter 3. There, I defended what I called Motivational Ought Implies Can (or MOIC) by appealing to the idea that if considerations that aren't possible motivators could give ultimate reasons, then ethics would no longer be in the business of guiding actions for beings like us. Instead, ethics would be action-guiding primarily for ideal people who are a lot smarter than us.

Now consider a simple extension of MOIC, that every normative reason is a possible motivator. If the regress works and everything up to the lowest level is excluded from giving reasons, and if this more general version of MOIC is true, then nothing gives reasons at all. The train of exclusion reasoning places a constraint on reasons such that they can only be considerations at the bottom level, and our Ought Implies Can principle generates a different constraint restricting reasons to what's motivationally possible. But no consideration

can satisfy both constraints. Thus reason-giving powers have drained away into nothing. So no one who believes in normative reasons can accept both generalized MOIC and PORE.

Now, PORE is part objection to the value-first view, and MOIC is a premise in the argument for that view. Adherents of PORE will have to deny a premise from that argument, and they are free to chose MOIC as the one to give up. Still, it's safe to deny MOIC and allow that even if the rational *ought* doesn't always imply *can*, most of us still can be motivated by most of the things we should be. If bottom-level considerations gave all the reasons, then we can almost never be motivated by the things we should be because they're all too complex to appreciate. If we are nearly universally unable to act for the genuine reasons we have, then not only does ethics better serve some ideal class of smarter people, it can *only* serve such people—it will never be action guiding for the vast majority of us. I think that's just too implausible to accept.

Here's the other side of that coin, which is not as damning, but still unpalatable. If bottom-level considerations were the only things that gave reasons, then in addition to turning the truths of ethics into unappreciable gibberish for us, it also turns out that all our everyday claims about the reasons we have false. To say that the resort's pleasantness gives me a reason to go there would not merely be to say something strictly false but in the neighborhood of the truth. It would be to say something so far removed from the truth that "pleasantness" is not plausibly shorthand for the more complicated thing we really mean. We can't mean the more complicated thing at all.

So even if MOIC is false, favoring still drains away into nothing if the

thought behind PORE is true. If we're faced with a choice between giving up on reasons altogether and giving up on exclusion, the sensible option has to be giving up on exclusion. In other words, we should say that PORE and its more specific instances are false—value-makers are not sufficient on their own to generate the reasons we have.

This concludes the defense against Scanlon's redundancy argument. If value doesn't give reasons, then it certainly doesn't give ultimate reasons, and the value-first view would be false. But as we've seen in this chapter, both premises of the redundancy argument are suspect. Redundancy is surely not the only objection to the value-first view, but it's definitely the most commonly accepted. Undermining this worry places the value-first view on much surer footing.

To close out the dissertation, we'll now turn to another powerful objection—this time against MOIC.

Chapter 6

Giving Up Hume's Guillotine

In chapter 3 I claimed that we needn't worry about any challenges to the principle that *ought* implies *can* from Hume's view that you can't get an *ought* from an *is*, because Hume's idea should be dispensed with. In this chapter I make good on that claim by making a point about the difference between truth and relevance.[1]

Hume's Guillotine rests on a distinction between normative and descriptive sentences.[2] This is because Hume's Guillotine is an inference barrier, and inferences deal in sentences. So, to say that you can't get an 'ought' from an 'is' is really to say that you can't validly infer any normative sentences from any descriptive ones. Though it was long out of favor, the Guillotine has enjoyed a resurgence among technically-minded philosophers who hope to prove that some version of it is true and silence the doubters. Proving something in a formal language is not enough to vindicate the principle, though. To be a *bona*

---

[1]An earlier and slightly different version of this chapter is forthcoming in *Australasian Journal of Philosophy*. Thanks to Taylor and Francis Publishing for generous author's rights.

[2]Hume's Guillotine is known by many other names. I borrow this more vivid label from Max Black 1964.

fide Hume's Guillotine, a principle must be a guide for ethics as we actually do it in natural languages. That is, it needs to govern real-life normative assertion. I suggest this as a criterion of success on any attempt to defend the view. Without this kind of relevance, there is little reason to care about Hume's Guillotine.

Because of powerful counterexamples from A.N. Prior, anyone defending the Guillotine must give an account of which sentences are normative and which are descriptive. Getting this right is crucial. It's one thing to prove a result in a language, and it's quite another to prove a result in one that can serve as an appropriate meta-language for the ethics we actually do.

I'll argue that the prospects for proving an acceptable version of Hume's Guillotine are dim, because the maneuvering required to secure a true principle alienates it from real-life normative assertion. This is a reason to think the effort to defend the Guillotine with formal methods is failing. That project may ultimately succeed, but too little attention goes to creating an adequate model of ethics. I hope to show the cost of that neglect.

In what follows I'll show that each successful strategy for beating Prior's counterexamples involves a faulty distinction between normative and descriptive sentences. Irrelevance across the wide range of views strongly suggests that the problem lies with the distinction between *sentences*. Some perfectly natural distinctions may not be clean enough for philosophical use, so choosing the right distinction to stake a claim on makes all the difference.

Many have argued that the Guillotine is problematic and said no more. I want to make the additional point that much of what's desirable in it can be had at the level of words rather than sentences: no descriptive terms are synonymous with any normative terms. Call this view *Moore's*

Law because of its connection with the naturalistic fallacy. The problems for Hume's Guillotine arise because it's an inference barrier and therefore requires a clear distinction between sentences. Moore's Law separates the normative from the descriptive without needing that. The inference barrier is replaced by a semantic one, and a term-level distinction takes the place of the sentence-level one. Insofar as it's worth having a methodological constraint like the Guillotine, Moore's Law offers a nice replacement.

Given how often the Guillotine is invoked, it's clear that many do want such a rule for ethics. But if what I argue is right, those of us who treat the Guillotine as philosophically secure way to correct students and rebuke opponents will have to think again. Those who have hoped to get even more mileage out it by enlisting it to show some further result will have to re-work their arguments.[3] The good news for all of us is that Moore's Law can deliver these goods as well.

The specific kind of relevance I'm after is this: a guiding principle like the Guillotine should be a rule for a game that people are really playing. An inference barrier couched in a language we're not speaking is no more significant than a barrier to inferences we're not making, and a false principle governs nothing. Since the versions of the Guillotine on offer require faulty ideas about what counts as normative, they're rules for the wrong game.

My concern with relevance is not the only one out there, however. Gerhard Schurz and Charles Pigden have engaged in a dispute over whether or not Hume's Guillotine is relevant in the sense of having other metaethical implications, particularly whether or not it implies anything

---

[3]Sinnott-Armstrong 2000, and Huemer 2005 invoke the Guillotine to defend views about moral epistemology, for example.

about non-cognitivism or naturalism.[4] This is a different question. Someone like Pigden who believes that it's metaethically neutral could also believe consistently that it's a guiding rule for ethics and ethical inferences. Pigden suggests he does think this when he says that people trying to infer normative conclusions from descriptive premises are making a significant mistake.[5] So metaethical relevance is not the same thing as first-order ethical relevance.

Although I doubt that the sentence-level distinction can be made well enough to deliver a relevant version of the Guillotine, I think there is a real difference and that some sentences are clearly normative or descriptive. That's compatible with thinking that explaining the difference is hard to do, and it's also consistent with allowing difficult borderline cases. Indeed, the basic problem that motivates the views we'll see below is brought out by such a case.

In the next section, I'll explain the dilemma for Hume's Guillotine that forces its defenders to give specific accounts of normative sentences. In sections three through five, I argue that efforts to frame an inference barrier around these accounts fail to produce a suitable version of the view. I conclude in section six by outlining how Moore's Law can stand in for the Guillotine, and offer at least one improvement on it. Just so, Moore's Law is well-positioned to be the guiding principle we wanted.

6.1 Prior's dilemma

A.N. Prior thought he could show Hume's Guillotine was false in a way that most other attempts could not. A popular strategy for arguing against the view went like this: identify a normative sentence, then argue that it follows

---

[4]See Pigden 2010b, ch. 6.

[5]Pigden 2010a, p. 26.

from some set of descriptive sentences. This was John Searle's approach in his example where Jones promises to to pay Smith five dollars.[6] From the descriptive premise that Jones uttered the words of the promise we can supposedly infer the normative conclusion that Jones must pay Smith five dollars. A cost of this strategy is that its success depends on whether the sentences really are normative and descriptive as alleged, and critics of Searle were quick to point this out. Prior's argument is especially powerful because it doesn't require consensus on which its crucial sentence is.

Start by considering this basic statement of the Guillotine.

Naïve HG: There are no satisfiable valid arguments from descriptive premises to normative conclusions.[7]

The argument against Naïve HG then goes as follows.[8]   Take the valid argument, A.

(**A**-i) Tea-drinking is common in England.

∴(**A**-ii) Either tea-drinking is common in England or all New Zealanders ought to be shot.

The premise of A is descriptive, so if the conclusion is normative then Naïve HG is false. But since it need not be true that all New Zealanders ought to be shot for (A-ii) to be true, one might suspect it's not normative. Suppose then that it's descriptive, and consider a second valid argument, B.

(**B**-i) It's not the case that tea-drinking is common in England.

---

[6]Searle 1964.   Other instances of the strategy can be found in Max Black 1964, Hector-Neri Castañeda 1973, and Peter Geach 1977.

[7]Being satisfiable means the premises are consistent. After all, anything follows from a contradiction.

[8]Prior 1960.

(**B**-ii) Either tea-drinking is common in England or all New Zealanders ought
     to be shot.

∴(**B**-iii) All New Zealanders ought to be shot.

Again, the premise is clearly descriptive. And plausibly, the tea-drinking
disjunction must be either normative or descriptive. If we say it's normative,
then argument A is a counterexample to Naïve HG. But if we say it's
descriptive, then B is a counterexample. Hence the dilemma.

The challenge for defenders of Hume's Guillotine is to get around Prior's
dilemma. By my count, there are three ways of doing this.

(1) Accept that A is a valid is-ought inference, but replace Naïve HG with
     something that doesn't apply to A (the term approach).

(2) Call the tea-drinking disjunction normative as a premise but descriptive
     as a conclusion (the relativity approach).

(3) Say that the tea-drinking sentence is neither normative nor descriptive
     (the fragility approach).

I'll discuss these strategies in order below.

6.2   First approach: terms

Normative sentences might be distinguished from descriptive ones by their
terms. Most words are clearly one or the other. On this strategy, a sentence
is normative just in case it contains a normative term, like 'ought', 'should',
'may', 'good', 'right', and 'wrong', and descriptive just in case it doesn't. Call
this the *term approach* to normative sentences.

The term approach handles the B-argument of Prior's dilemma where the disjunction is a premise, because it has an 'ought'. But it doesn't block the A-argument where the disjunction is the conclusion, since its premise has no normative terms. Yet Frank Jackson observes that there is something odd about the A-argument: the 'ought to be shot' in the conclusion is replaceable by any grammatical expression whatsoever—'are friendly' or 'prefer rugby to curling'—and the argument remains valid, since it's an instance of the rule of addition. Call a term in an argument *vacuous* just in case each occurrence is uniformly replaceable without prejudice to the inference's validity.[9] This insight gives way to a version of Hume's Guillotine that avoids the looming counterexample.

Jackson's principle rules out any argument with vacuous terms from the scope of the view.

JHG: No argument from premises lacking normative terms to a conclusion with normative terms is valid, unless one of its terms is vacuous.[10]

Prior's A-argument is no counterexample to this version of Hume's Guillotine, since the 'ought' of the tea-drinking disjunction is vacuous. But the restriction against any vacuous term goes too far. Jackson's principle places any argument that is valid in virtue of its logical form outside of its scope. Think of modus ponens for example: in an inference of the form *if p then q, p; therefore q* both *p* and *q* are vacuous. Thus the only arguments to which JHG properly applies are those that are valid just in virtue of the meaning of their terms.[11]

It's not difficult to see why this goes too far. Philosophical ethics is in the business of giving formally valid arguments, as well as informal ones. So if JHG

---

[9]c.f. Pigden 1989, p. 134.

[10]Jackson 1974, p. 93.

[11]See Jackson 1974, p. 91.

has nothing to say about arguments that follow modus ponens, disjunctive syllogism, etc., then it cannot be a guiding principle of ethics as we actually do it.

Charles Pigden extends Jackson's strategy by framing a version of the Guillotine around vacuity, but only when it occurs in the conclusion. Roughly, Pigden's says:

PHG: No argument from premises lacking normative terms to a conclusion with normative terms is valid, unless the normative terms in the conclusion appear vacuously.

One immediate concern about PHG (that also affects JHG) is that it's under-specified—it fails to distinguish use from mention. A sentence might contain a term in either way, so the term approach really covers two distinct ideas. There is a *narrow* claim which says that a sentence is normative just in case it uses a normative term, and a *wide* one which says that a sentence is normative just in case it uses or mentions a normative term. I'll tackle the narrow version first.

NHG: No argument from premises not using normative terms to a conclusion that uses normative terms is valid, unless the normative terms in the conclusion appear vacuously.

I believe this is what Pigden had in mind. Pigden does not explicitly consider the use-mention distinction, but he speaks of predicates occurring, not individual constants, so it's safe to assume this is the principle he would choose.[12]

---

[12]Pigden 1989, p. 136.

The trouble with NHG is that while it avoids Prior's dilemma, it's susceptible to another kind of counterexample. To see how the example works, consider this wholly descriptive argument.

(**C**-i) 'Clark Kent' names the reporter from Krypton who writes for the Daily Planet.

∴(**C**-ii) Clark Kent writes for The Daily Planet.

Argument C is truth-functionally valid in the sense that the truth of the premise guarantees the truth of the conclusion. It's not formally valid, but restricting the Guillotine to just formally valid arguments won't do, as we've seen. And although the conclusion may not be derivable by syntactic inference rules, it does clearly follow from the naming convention established by the premise. Once it's established, we are licensed to use the name. Furthermore, 'Clark Kent' is clearly non-vacuous. Replacing the conclusion with 'Lois Lane writes for the The Daily Planet' would invalidate the argument. Now consider argument D, which trades on the same kind of inference.

(**D**-i) 'Ought to' names the relation that holds between Lois and the action-type *donate to charity.*

∴(**D**-ii) Lois ought to donate to charity.

If you're thinking that the premise of this argument is normative, I'm inclined to agree. It's just a roundabout way of saying that Lois ought to donate to charity. Absent a particular view about makes for a normative sentence, the D-argument is not an especially strong candidate for an is-ought derivation.[13]

But responding to Prior's dilemma requires particular criteria for normative sentences because Naïve HG looks false. For NHG, it's the narrow term

---

[13]I imagine this is why no one else I know of has tried this sort of counterexample.

approach. It says (D-i) is descriptive, (D-ii) is normative, and the 'ought' in the conclusion is non-vacuous. So D is a counterexample to NHG, provided the argument is valid. And as C shows, it's not a contrived one-off argument either. It's not hard to imagine genuine instances of this kind of reasoning. Perhaps an enemy of Superman has obtained a government list of known superhero identities, hoping to threaten those Superman cares about. Or perhaps a powerful computer had generated a list of possible actions for Lois that carry extremely high expected utility, but due to limitations in the programming language, results are given metalinguistically. Given the information in the premise the inference to the conclusion is an easy one, but that doesn't make D any less of a counterexample.

Defenders of Pigden will be quick to point out that he has given a proof, so there must be something amiss with the counterexample. But the trouble really lies with his proof. The problem is that it can't show NHG because it fails to distinguish use from mention. Pigden hopes to prove that a predicate "cannot occur non-vacuously in the conclusion of a valid inference unless it appears among the premises."[14] Again, predicates can occur in both ways. Argument D reveals that the proof shows at most that a predicate cannot be used in the conclusion of a valid argument unless it is used or mentioned. But that's not the same as demonstrating NHG, that's showing something about the wide term view, which is weaker—the disjunction is crucial.

Objectors might try another response. In a recent commentary on different counterexamples—which infer normative conclusions from descriptive premises by claiming some group of sentences is true and including some normative-looking sentence among them—Pigden objects that,

---

[14]Pigden 1989, p. 136.

"Both...can be accused of concealing the moral content of one of their premises behind the truth predicate."[15]

Likewise, I can be accused of concealing the moral content of the premise in the D-argument behind the naming convention it establishes, and I readily admit that I have. The question is whether my bit of linguistic trickery makes a difference to NHG. In his commentary, Pigden speaks of a distinction between "formally" moral and "substantively" moral sentences.[16] Some may therefore object that even if the premise of D is formally descriptive by the rules of NHG, it is nonetheless substantially normative and so fails to undermine the spirit of the Guillotine.

Yet this sort of objection gets things backwards. Adherents of strategy 1 for combating Prior's dilemma are not entitled to fall back on some more general or "substantial" version of the view if the formal candidate fails. Since the predicate approach allows Prior's tea-drinking disjunction as normative, Naïve HG is false by its lights. It's this trouble with the naïve claim that forces defenders of the Guillotine to come up with something more sophisticated. If one of the sophisticated principles succeeds, that just is the substantial view. If it fails, it's time to go back to the drawing board, since there is nothing defensible to fall back on. Those who want to defend Hume's Guillotine cannot have it both ways. Either the replacement for Naïve HG is the genuine article, or it isn't.

The D-argument also undermines Schurz's defense of the Guillotine, which categorizes sentences as either purely normative, purely descriptive, or mixed. Because he wants to prove versions of it for a range of deontic logics, Schurz identifies sentences by deontic operators rather than logical constants, but the

---

[15]Pigden 2010a, p. 36.

[16]Pigden 2010a, pp. 34-38.

basic approach is the same as Pigden and Jackson's: the words themselves make difference. A sentence is purely normative just when the entire thing lies inside the scope of a normative operator, purely descriptive just when there is no use of a normative operator, and mixed just when the operator binds only part of the sentence.[17] So a sentence like 'Lois ought to donate to charity' will be purely normative, and a sentence like 'snow is white' will be purely descriptive. Because only its right-hand disjunct is in the scope of a normative operator, Prior's tea-drinking disjunction comes out mixed.

Given these categories, there are three possibilities for a valid is-ought inference. We might have (a) a purely descriptive to mixed (like Prior's A-argument where the disjunction is the conclusion), (b) mixed to purely normative (like Prior's B-argument where the disjunction is a premise), or (c) purely descriptive premises to purely normative. Schurz appears to think others have proved that no non-trivial type (c) arguments are valid, so he leaves this possibility out and focuses on blocking types (a) and (b).[18]

But ignoring (c) is a mistake—the D-argument constitutes a valid instance by Schurz's lights. The naming convention for Lois given by the premise must be interpreted as purely descriptive here. Although there is a mention of 'ought' and operators are defined syntactically, it can hardly be thought of as an active normative operator. Trying to interpret the premise that way leads to nonsense, since there would be no grammatical sentence left to operate on ('O' stands for the operator, with the sentence it binds in brackets):

O[names the relation between Lois and the action-type *donate to charity*]

This is not a formula of any language. It's a misuse of the sentential operator, so clearly it won't do. The premise is therefore purely descriptive. And the

---

[17]Schurz 2010, pp. 200-201.

[18]See Schurz 2010, p. 204, and Schurz 1997, ch. 3.3 for discussion.

conclusion is purely normative, since there is an active normative operator binding the whole sentence:

$$O[\text{Lois donates to charity}]$$

So although Schurz's view differs from NHG, it fails for the same reasons.

The failure of the narrow approach drives us to the wide one. It has no trouble handling the D-argument because it designates any premise that uses or mentions a normative term as a normative sentence. Applying it to Pigden's PHG gives us:

WHG: No argument from premises that do not use or mention normative terms to a conclusion that uses or mentions normative terms is valid, unless the normative terms in the conclusion appear vacuously.

This principle is much more secure: the D-argument can't be a counterexample because it mentions 'ought', and Pigden has a proof of it. The issue with WHG is not falsity, but irrelevance. If WHG is going to be the Guillotine, wide term approach to normative sentences needs to be adequate. But consider:

(5L)   The word 'ought' has five letters in it.

The wide approach incorrectly calls this a normative sentence. WHG is true, but pairing it with the wide approach means it's a Guillotine for a language no one speaks. In English, 5L is just a piece of orthography—even if it's normative in some sense, it's not the sort of moral claim that's relevant here. If merely describing a norm like the spelling of 'ought' makes a sentence normative for

the purposes of the Guillotine, then Prior's dilemma wouldn't have been a problem in the first place, since the tea-drinking claim describes a norm too.

Some might counter that admitting too much into the set of normative sentences isn't a problem. As long as WHG governs all the sentences we need it to, maybe it's acceptable for it to govern some that weren't intended. I agree that it's not a problem if there's reason to group the unintended cases with the original ones. To draw an analogy, it doesn't count against laws meant regulate telegraph transmissions that they also rule over telephone and internet traffic. But in this case there are good reasons against grouping sentences like 5L with the normative ones. Grammar and linguistics are not ethics, nor anything like it. To see this, notice that being a skeptic about ethics gives us no reason to doubt grammar.

So, neither version of the term approach delivers Hume's Guillotine.

6.3   Second approach: relativity

A different approach to Prior's dilemma is to deny that either of his arguments can be sound counterexamples. Toomas Karmo suggests that a sentence is normative or descriptive relative to a world. Worlds are composed of three maximal consistent sets: a set of obviously normative sentences (like 'Lois ought to donate to charity'), a set of obviously descriptive ones (like 'snow is white'), and a set of non-obvious sentences (like 'snow is white or Lois ought to donate to charity'). A non-obvious sentence is normative just in case changing the truth-value of some obviously normative one, while holding the obviously descriptive ones fixed, changes its truth-value. So, for a non-obvious case like 'snow is white or Lois ought to donate to charity', its status will depend on

the world of evaluation. If snow is actually white, then varying obviously normative sentences won't change the truth-value because the first disjunct is fixed. But if snow isn't white at the world of evaluation, then the truth-value depends on the obviously normative sentences because the second disjunct is in that set. Call this the *relativity approach* to normative sentences.[19]

This view allows Karmo to prove a candidate for the Guillotine that avoids Prior's dilemma.

KHG: There is no argument whose premises are descriptive and whose conclusion is normative, relative to worlds where it's sound.[20]

Given the relativity approach, KHG avoids Prior's dilemma because the crucial tea-drinking disjunction varies between the two arguments. It's descriptive as the conclusion of A because its truth is preserved by the premise: it's true regardless of what ought to happen to New Zealanders. As a premise in B it's normative. Its truth depends on normative facts about Kiwis because B is sound only if tea-drinking is scarce.

Some normative-sounding assertions create a problem for the relativity approach. Consider the following:

(BILL)   Bill was right to tell the truth about Monica.

To know whether BILL is normative or descriptive, we need to specify a world. Start with a world where Bill lies.

---

[19]Karmo's view draws on I.L. Humberstone 1982, which is inspired in turn by J.M. Shorter's 1961 reply to Prior.

[20]Karmo 1988, p. 256, c.f. Humberstone 1982.

Since BILL is a factive claim about satisfying a norm, it's equivalent to a conjunction: *Bill told the truth about Monica and Bill ought to tell the truth about Monica.* We'll run Karmo's test on this for simplicity. If varying an obviously descriptive sentence while fixing the descriptive ones forces it to change truth-value, we know it's normative. But since its first conjunct is descriptive and false, that won't happen. The conjunction is descriptive when Bill lies, so BILL is too.

That BILL is descriptive at any world where he lies already invites a charge of irrelevance because it means KHG is a rule for the wrong object language. It leaves out sentences that should be in the scope of the view. And the divergence from natural language gets worse. Notice that while BILL is descriptive at the actual world, it is normative according to the view at any world where Bill does tell the truth. Then consider:

(BILL*)   Bill was right *not* to tell the truth about Monica.

Parallel reasons show BILL* is normative anywhere Bill doesn't tell the truth. But there is no normatively significant difference between the two—each makes a clear moral evaluation. The only difference is that at some worlds the sentences correctly describe Bill's action and in others they don't.[21] Yet it's difficult to see how this would be relevant to assessing normativity. If it isn't relevant, Karmo's approach doesn't accurately model natural language.

Some might argue that correctly describing Bill's action is normatively

---

[21]I've heard it suggested that sentences with presupposition failure lack truth-values. I am skeptical that this is the case with BILL, but responding this way won't help HGK. First, if BILL has no truth-value, it cannot be normative on Karmo's definition. Second, since BILL appears to imply truth-apt sentences like 'Bill told the truth', there's a further problem of explaining how we can get a sentence with a truth-value out of one without. People interested in putting up inference barriers might not want to allow this. Thanks to an anonymous referee for the second point.

relevant by comparing these cases with Prior's disjunction. Because the disjunction is descriptive according to the view when it describes the facts about tea-drinking correctly and normative when it doesn't, it gets a mixed treatment, like BILL and BILL*. If it's acceptable for Prior's disjunction to vary with correctness, then perhaps it really is relevant to whether a sentence is normative.

Yet even if we accept the mixed treatment of Prior's disjunction—and we needn't—that would show that correct description is normatively relevant only if correctness does some work toward explaining why we accept different verdicts. Otherwise correctness might have nothing to do with normativity. Other explanations are plausible: the mixed treatment of Prior's disjunction is tolerable because of what asserting it would commit us to at different worlds. At worlds where we know that tea-drinking is common in England, we can assert the disjunction while denying that New Zealanders ought to be shot. But when we consider worlds where we know that tea-drinking is not common, asserting the disjunction commits us to saying that all New Zealanders should, in fact, be shot. Karmo's relativity approach reflects the fact that at some worlds we would be committed to obviously normative claims, and not at others.

Notice there is no similar change in our commitments when we assert BILL or BILL*. Whatever the world, saying that Bill was right to tell the truth about Monica means that Bill ought to tell the truth about Monica. That's a reason for thinking at least some normative sentences stay that way across worlds. And if they do, then KHG governs the wrong language and it too is a poor candidate to be Hume's Guillotine.

6.4   Third strategy: fragility

Perhaps a different modal logic is the answer. Deontic logic is one where the concepts of permission and obligation are defined like possibility and necessity operators, but within a special set of morally satisfactory worlds. Roughly, a sentence is permissible just when it's true at at least one morally satisfactory world, and a sentence is obligatory just when it's true at all the morally satisfactory worlds.

In a deontic system, the truth of ought-sentences is sensitive to changes to the set of worlds that are morally satisfactory in a given model. Greg Restall and Gillian Russell use this feature to dodge to Prior's dilemma. Call it the *fragility approach.* It calls a sentence is normative just in case there is at least one model where replacements and additions to the set of satisfactory worlds change its truth-value—that is, one model in which it's fragile—and descriptive just in case it isn't sensitive to those replacements in every model where it's true, making it preserved.[22] A sentence like 'Lois ought to donate to charity' is normative on this approach because its truth does change with the set of satisfactory worlds: if Lois donates at all the satisfactory worlds, and we add a new satisfactory world to the model in which she doesn't, the sentence goes from true to false. Conversely, a sentence like 'snow is white' comes out descriptive because it doesn't matter what's in the set of satisfactory worlds. Since 'snow is white' is not a deontic modal claim, its truth-value doesn't vary with things happening at other possible worlds.

This approach is powerful enough that Restall and Russell don't need to state an alternative to Naïve HG. It's left untouched by Prior's arguments because the crucial tea-drinking disjunction is neither normative

---

[22]Restall and Russell 2010, pp. 254-255.

nor descriptive. It fails the test for normativity in any model where tea-drinking is actually common in England because it will remain true no matter what additions or replacements are made to the set of satisfactory worlds, so it's not fragile. If tea-drinking is actually scarce in England, then it fails the test for being descriptive.The disjunction would be true only if all New Zealanders are shot at every satisfactory world. Replacing one of these worlds with one at which they are not shot makes the disjunction false, so the disjunction is not preserved when true. Thus neither of Prior's arguments counts as an is-ought derivation.

Peter Vranas argues that the fragility approach mis-categorizes some normative sentences, making the view extensionally inadequate.[23] Consider:

(LC) Lois should donate to charity if she is able.

Vranas would say LC is equivalent to *either Lois is unable to donate or she ought to donate to charity*, which is labelled as non-normative by the fragility approach for the same reason as Prior's disjunction: it's not fragile in models where Lois is unable to donate. Vranas believes that's implausible.

This is the right line of argument to press against the fragility version of the Guillotine, but it's insufficient for present purposes. I'm suggesting that the approach cannot generate an inference barrier relevant to the actual practice of ethics. Vranas's argument doesn't show this for two reasons: (a) it fails to establish that fragility approach really is inadequate, and (b) as we saw with the wide term approach, inadequacy on its own may not lead to irrelevance. More needs to be said.

The argument doesn't show what it intends to because it says nothing about the scope of the obligation operator in LC. This is crucial. If it has wide

---

[23]Vranas 2010, pp. 264-265.

scope: $\Box(Al \Rightarrow Dc)$, then the sentence is not equivalent to the non-normative disjunction $\neg Al \vee \Box Dc$. A wide-scoped LC would come out as normative because in any model where it's true, adding or replacing satisfactory worlds can make the sentence false, making it fragile. Just add a world where some people who are able to donate refuse. The fragility approach is inadequate only if we can show that LC has narrow scope: $Al \Rightarrow \Box Dc$. Given LC's surface grammar, it's not obvious which reading is correct.[24] So rather than beginning with a sentence of English and insisting on narrow scope, a different strategy is needed.

I think the right approach is to start in the meta-language: take a narrow-scoped formula of deontic logic like $Al \Rightarrow \Box Dc$. Then we ask if it has any acceptable natural language translations that are normative.The most straightforward is this:

(LC*) If Lois is able, she ought to donate to charity.

To the extent the original LC is normative—and it is—LC* is too. That's why the fragility Guillotine is inadequate.

But mis-categorizing doesn't always guarantee irrelevance, so another step is needed. Remember the telegraph laws that wound up governing internet traffic. In that case, the wide term approach grouped non-normative sentences in with the normative ones. The fragility approach does the opposite by putting normative cases in with the non-normative, so it's is not a situation where all the sentences that need to be governed are within the scope of the principle. This barrier fails to govern what it should, so it's irrelevant, just as Jackson's version was for leaving out valid arguments that aren't formally so.

---

[24]Some people, notably John Broome 1999, think that sentences like LC should always be understood to have wide scope. I think that's a non-starter, since narrow scoped imperatives that look just like LC are easily derivable from $\neg Al$.

6.5 Moore's Law

The challenge Prior's dilemma creates looks difficult to surmount. The term, relativity, and fragility approaches each generate a true principle that avoids the counterexamples, but every one is irrelevant to actual normative assertion. This suggests there may be no candidate that is both true and adequate. Others may prove additional candidates that look like the Guillotine, but given the difficulties above we should be hesitant to accept them as the genuine article without thorough scrutiny.

The general problem comes from attempting frame a philosophically significant inference barrier around the distinction between normative and descriptive *sentences*, which is difficult to pin down. Moore's Law steers clear of these problems because it's a semantic barrier: no atomic normative *terms* are synonymous with any atomic descriptive terms, either directly or by substitution. I think Moore's Law can both stand in for the Guillotine and improve on it in an important way. I won't argue for Moore's Law, though I do think the open question argument offers a good reason to accept it. What I will try to show instead is that Moore's Law can deliver a lot of the regulation of ethics that Hume's Guillotine promised.

Of course, Moore's Law is not without its own critics. Some metaethical naturalists think it shows very little. After all, 'the morning star' and 'the evening star' have different meanings and are conceptually independent, but co-refer all the same. So too could normative terms be undefinable by descriptive terms, but refer to the same properties. But this is more a criticism of what we can learn from Moore's Law than it is an indictment of the principle itself. The only view Moore's Law rules out is analytic naturalism—'goodness' and 'pleasure' can't mean the same thing, for instance. Those who aren't

analytic naturalists are free to accept Moore's Law and everything it implies. Which properties are identical, and thus which expressions are co-referring, is another question entirely.

Some may doubt the principle's suitability for other reasons. If we are called upon to give a precise account of the distinction between normative and descriptive terms, we risk making it a rule for the wrong game too. Recall, though, that Hume's Guillotine needs to be that careful only because of Prior's dilemma. Had Prior not backed it into a corner, the barrier could have stood on a more rough and ready version of the distinction. I think the same is true of Moore's Law: provided there is no equivalent problem, it should to be enough to specify a list of normative terms—like 'ought', 'should', 'may', 'good', 'right', and 'wrong'—that contrast with descriptive ones and leave the question of how to specify the difference between them for another day.

So is there an analogue of Prior's dilemma to make trouble for Moore's Law? Creating one requires choosing a term and asking if it's normative or descriptive. If the moorean says it's descriptive, then the objector shows that it's synonymous with a normative term. If it's said to be normative, tho objector shows it has the same meaning as a descriptive term.

At first glance it's difficult to see how to get this started. To follow Prior we would want to find something disjunctive that has both normative and descriptive elements, like 'quick or wrong'. The trouble is that Moore's barrier is meant to govern atomic terms only, so disjunctive predicates look beyond the scope of the Law. Things aren't so simple, however. Predicate abstraction in logic enables us to represent seemingly compound expressions as atomic predicates of a more precise meta-language. Phrases like 'x is quick or x is wrong' can be formalized into a unitary constant that takes x as its

argument. That is, x satisfies the one-place predicate *quick-or-wrong*, which we understand as Qx—an atomic term.

So far so good for the opponent of Moore's Law, allowing that predicate abstraction gives an adequate model for natural language. The next step would be to show that our new single expression means the same as both normative and descriptive terms, just as Prior's tea-drinking disjunction participates in valid implication with normative and descriptive sentences. Here is where the analogy to Prior breaks down. What would the quick-or-wrong predicate be synonymous with? I don't believe it could be either 'quick' or 'wrong' on its own, as each represents just a third of the complex. Failing those, no other candidates present themselves. More generally, it's difficult to imagine a clearly normative or clearly descriptive term of natural language with a disjunctive meaning comparable to "quick-or-wrong" or anything like it. Sure, we could introduce a new arbitrary term and stipulate that it has such a meaning, but that won't help because it creates a new object language. Moore's Law governs only natural languages. Creating a dilemma in an artificial one is irrelevant.

There may be a sense in which 'quick' and 'wrong' are partially synonymous with the complex predicate. Each does capture part of the meaning. But this won't help either, since objecting that some normative terms are partially synonymous with a descriptive term (or vice versa) would be like objecting to Hume's Guillotine by suggesting that some descriptive sentence partially implies a normative conclusion because it could be used in an argument for that conclusion. To get a genuine objection we need complete valid implication and complete synonymy.

Prior had disjunctive syllogism and the rule of addition at his disposal, which allowed him to piece together valid inferences with the tea-drinking

sentence and its disjuncts. There is no equivalent way to break up a unitary predicate since it has no proper parts to work with. The move that was needed to state the dilemma prevents us from making it work. That's a good reason to think there is no analogue to Prior's argument here.

With those worries out of the way, all that remains is to show why friends of the Guillotine can accept Moore's Law as a substitute. Many present-day advocates should already accept Moore's Law. Karmo's relativity principle assumes Moore's Law because synonymy between normative and descriptive sentences would make it impossible to vary the set of obviously normative sentences while fixing the obviously descriptive sentences. Restall and Russell's fragility barrier also assumes the Law, since it treats descriptive terms as predicates and normative terms as deontic operators. As a rule, atomic predicative sentences imply deontic modal sentences only in certain models. But if some normative terms had the same meaning as descriptive ones, then some atomic sentences would imply deontic modals in every model, and that can't be. Pigden has also expressed support for Moore's Law, though he thinks the Guillotine doesn't imply anything about it.[25]

Some who defend the Guillotine have an intrinsic interest in it because they believe it's an instance of the conservativeness of logic: you can't get out what you haven't put in. But many more have been motivated by its ability to guide ethics. I have my doubts about certain interpretations of conservativeness because of cases like the Clark Kent D-argument, but if Moore's Law can regulate ethics in the other ways the Guillotine was meant to, this demand can be met.

The Guillotine is employed primarily to explain why certain arguments

---

[25]Pigden 1989, p. 128.

are invalid, like *everyone does x, therefore x is permissible.* Hume himself
appealed to the Guillotine to show that inferences like *God favors x, therefore
x ought to be the case* and *x is human nature, therefore x is permissible* are
invalid.[26] But Moore's Law can explain this too, even if Hume himself might
not have accepted it. These inferences are not formally valid, so if they were
it would likely be in virtue of meaning. But if Moore's Law is true, there are
no normative-descriptive synonyms, so arguments like these fail because there
is no chain of substitution of synonyms for synonyms to license them. This
explanation improves on the one from the Guillotine. It preserves a sense in
which you can't get an 'ought' from an 'is', it does more than simply assert
that the inference can't be made, and it avoids the problematic sentence-level
distinction.

This might seem too quick, though. Moore's Law denies that normative
and descriptive terms mean the same, but failure of synonymy doesn't always
result in an inference barrier. The word 'uncle' isn't synonymous with 'male',
but the fact that Clark is an uncle validly implies that Clark is male. So
perhaps Moore's principle can't block the arguments that the Guillotine was
meant to after all.

While it's true that some inferences cross gaps in meaning, there is
an important difference between the uncle-to-male argument and the ones
Hume had in mind. In the uncle-male case, synonymy does secure the
conclusion—not directly, but by substitution. To derive 'Clark is male' from
'Clark is an uncle' we replace 'uncle' with 'father's brother', then 'brother'
with 'male sibling', and we have the conclusion by simplification. The
human-nature-to-permissible argument (and others like it) cannot work this

---

[26]Hume 1739/1978, p. 469.

way if Moore was right. If it did, there would have to come a point in the substitution chain where a descriptive term was replaced with a normative one on the basis of meaning. Moore's Law says there is a divide that cannot be bridged—the only terms that can stand in for 'human nature' are descriptive ones, and the only ones that can stand in for 'permissible' are normative ones.

Others have relied on the Guillotine to argue that normative beliefs cannot be deductively justified by descriptive beliefs, like Huemer and Sinnott-Armstrong.[27] Moore's Law cannot show this of course, since beliefs are propositional and Moore's Law says nothing about sentence-level things. This is for the best though, because the same problems that make trouble for the inference barrier will afflict a barrier between normative and descriptive beliefs. Yet Moore's Law can show something similar: no beliefs featuring normative terms can be justified by deduction from a set of premises lacking normative terms by substitution of synonyms for synonyms. Moore's barrier shows that the chain of substitution cannot be completed.

Still others have tried to leverage the Guillotine to create an explanatory advantage for moral non-cognitivism, the idea that normative assertions lack truth-values.[28] Non-cognitivists have an easy explanation of why Hume's Guillotine would be true—since normative sentences have no truth-values, they cannot be implied by any sentences whatsoever—while cognitivists are supposed to find it difficult to offer a competing explanation. I take no stand on whether non-cognitivists would have such an advantage, but to the extent that there would be one, Moore's Law can be employed for the same work. Non-cognitivists can instead explain the failure of synonymy between sentences using normative and descriptive terms by appealing to the former's

---

[27]Sinnott-Armstrong 2000; Huemer 2005.
[28]See Hare 1952 in particular, and many of the entries in Pigden, ed. 2010.

lack of truth-aptness, and claim as before that cognitivists cannot offer a good competitor.

There may yet be a way to rescue Hume's Guillotine from Prior's dilemma, but the current prospects are dim, since the troubles with normative and descriptive sentences cast doubt on the enterprise of proving it. If what I've argued about Moore's Law is correct, then the heavy lifting needed to save the Guillotine from the mire of problems that emerged above is unnecessary. Giving it up involves little if any cost and affords at least one important additional benefit. Devotees would therefore do better to focus on defending and invoking Moore's Law instead.

References

[1] Anderson, Elizabeth 1993. *Value in Ethics and Economics* (Cambridge: Harvard University Press).

[2] Anscombe, G.E.M. 1957/2000. *Intention* (Cambridge: Harvard University Press).

[3] Aquinas, Thomas 1948. *Summa Theologica*, trans. Fathers of the English Dominican Province (New York: Benziger Brothers).

[4] Aristotle. *Nicomachean Ethics*, trans. J.A.K. Thompson (New York: Penguin Classics).

[5] Benacerraf, Paul 1973. "Mathematical Truth." *Journal of Philosophy* 70: 661–679.

[6] Black, Max 1964. "The Gap Between 'Is' and 'Should'." *Philosophical Review* 73: 165–181.

[7] Block, Ned 2003. "Do Causal Powers Drain Away?" *Philosophy and Phenomenological Research* 67:133–150.

[8] Brentano, Franz 1889/1902. *The Origin of the Knowledge of Right and Wrong* (London: Archibald Constable).

[9] Broome, John 1999. "Normative Requirements." *Ratio (new series)* 12: 398–419.

[10] Brown, Campbell 2011. "Consequentialize This." *Ethics* 121: 749–771.

[11] Castañeda, Hector-Neri 1973. "On the Conceptual Autonomy of Morality." *Noûs* 7: 67–77.

[12] Chalmers, David 2011. "Verbal Disputes." *Philosophical Review* 120: 515–566.

[13] Churchland, P. M. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78: 67–90.

[14] Crisp, Roger 2005. "Value, Reasons, and the Structure of Justification: How to Avoid Passing the Buck." *Analysis* 65: 80–85.

[15] —— 2008. "Goodness and Reasons: Accentuating the Negative." *Mind* 117: 257–265.

[16] Dancy, Jonathan 2000. "Should We Pass the Buck?" *Royal Institute of Philosophy Supplement* 47: 159–173.

[17] —— 2002. *Practical Reality* (New York: Oxford University Press).

[18] D'Arms, Justin and Jacobson, Daniel 2000. "Sentiment and Value." *Ethics* 110: 722–748.

[19] Davidson, Donald 1970a. "How Is Weakness of the Will Possible." In *Moral Concepts* Joel Feinberg ed. (New York: Oxford University Press).

[20] —— 1970b. "Mental Events." In *Experience and Theory*, Forester and Swanson eds. (Humanities Press): 79–101.

[21] Dreier, Jamie 1993. "Structures of Normative Theories." *The Monist* 76: 22–40.

[22] Enoch, David 2011. *Taking Morality Seriously: A Defense of Robust Realism* (New York: Oxford University Press).

[23] Ewing, A.C. 1948. *The Definition of Good* (London: Hyperion Press).

[24] Fodor, Jerry 1989. "Making Mind Matter More." *Philosophical Topics* 17: 59–79.

[25] Geach, P.T. 1977. "Again the Logic of 'Ought'." *Philosophy* 52: 473–476.

[26] Hare, R.M. 1952. *The Language of Morals* (Oxford: Clarendon Press).

[27] Hooker, Brad and Stratton-Lake, Philip 2006. "Scanlon Versus Moore on Goodness." In *Metaethics After Moore*, Terry Horgan and Mark Timmons eds. (Oxford: Oxford University Press).

[28] Howard-Snyder, Frances 2006. "'Cannot' Implies 'Not Ought'." *Philosophical Studies* 130: 233–246.

[29] Huemer, Michael 2005. *Ethical Intuitionism* (London: Palgrave Macmillan).

[30] Humberstone, I.L. 1982. "First Steps in Philosophical Taxonomy." *Canadian Journal of Philosophy* 12: 476–478.

[31] Hume, David 1739/1978. *A Treatise of Human Nature*, eds. P.H. Nidditch and L.A. Selby-Bigge, (New York: Oxford University Press).

[32] Jackson, Frank 1974. "Defining the Autonomy of Ethics." *Philosophical Review* 83: 88–96.

[33] Kagan, Shelly 1989. *The Limits of Morality* (New York: Oxford University Press).

[34] Karmo, Toomas 1988. "Some Valid (But No Sound) Arguments Trivially Span the 'Is'-'Ought' Gap." *Mind* 97: 252–257.

[35] Kim, Jaegwon 1992. "The Nonreductivist's Troubles With Mental Causation." In *Mental Causation*, John Heil and Alfred R. Mele eds. (New York: Oxford University Press).

[36] Kraut, Richard 2014. "Aristotle's Ethics." In *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), Edward N. Zalta ed., URL = <http://plato.stanford.edu/archives/sum2014/entries/aristotle-ethics/>.

[37] Lang, Gerald 2008. "The Right Kind of Solution to the Wrong Kind of Reason Problem." *Utilitas* 20: 472–489.

[38] Liao, S. Matthew 2010. "The Buck-Passing Account of Value: Lessons From Crisp." *Philosophical Studies* 151: 421–432.

[39] Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong* (New York: Penguin).

[40] Menzies, Peter 2004. "Difference-Making in Context." In *Causation and Counterfactuals* J. Collins, N. Hall, and L. Paul eds. (Cambridge: MIT Press): 139–179.

[41] Merricks, Trenton 2001. *Objects and Persons* (New York: Oxford University Press).

[42] Moore, G.E 1903. *Principia Ethica* (Cambridge: Cambridge University Press).

[43] —— 1942. "A Reply to My Critics." In *The Philosophy of G. E. Moore*, ed. Paul A. Schilpp (Evanston: Northwestern University Press): 533–677.

[44] Nolan, Daniel 1997. "Impossible Worlds: A Modest Approach." *Notre Dame Journal of Formal Logic* 38: 535–572.

[45] Olson, Jonas 2006. "G.E. Moore on Goodness and Reasons." *Australasian Journal of Philosophy* 84: 525–534.

[46] Parfit, Derek 2011. *On What Matters* (Oxford: Oxford University Press).

[47] Pigden, Charles 1989. "Logic and the Autonomy of Ethics." *Australasian Journal of Philosophy* 67: 127–151.

[48] —— 2010a. "Introduction." In *Hume, 'Is', and 'Ought'*, ed. Pigden (London: Palgrave Macmillan): 1–38.

[49] —— 2010b. "On the Triviality of Hume's Law: A Reply to Gerhard Schurz." In *Hume, 'Is', and 'Ought'*, ed. Pigden (London: Palgrave Macmillan): 217–238.

[50] Plato 1997a. "Meno." In *Plato: Complete Works*, Cooper and Hutchinson eds. (Indianapolis: Hackett).

[51] —— 1997b. "Gorgias." In *Plato: Complete Works*, Cooper and Hutchinson eds. (Indianapolis: Hackett).

[52] Prior, A.N. 1960. "The Autonomy of Ethics." *Australasian Journal of Philosophy* 38: 199–206.

[53] Raz, Joseph 2010. "On The Guise of the Good." In *Desire, Practical Reason, and the Good*, ed. S. Tenenbaum (Oxford: Oxford University Press): 111–137.

[54] Restall, Greg and Russell, Gillian 2010. "Barriers to Implication." In *Hume, 'Is', and 'Ought'*, ed. Pigden (London: Palgrave Macmillan)

[55] Scanlon, T.M. 1998. *What We Owe to Each Other* (Cambridge: Harvard University Press).

[56] ―――― 2002. "Replies to Wallace, Dworkin, and Deigh." *Ethics* 112: 507–528.

[57] Schroeder, Mark 2011. "Buck-Passers' Negative Thesis." *Philosophical Explorations* 12: 341–347.

[58] Schurz, Gerhard 1997. *The Is-Ought Problem: An Investigation in Philosophical Logic* (Dordrecht: Kluwer Publishing).

[59] ―――― 2010. "Non-trivial Versions of Hume's Is-Ought Thesis." In *Hume, 'Is', and 'Ought'*, ed. Pigden (London: Palgrave Macmillan).

[60] Searle, John 1964. "How to Derive 'Ought' from 'Is'." *Philosophical Review* 73: 43–58.

[61] Shafer-Landau, Russ 2003. *Moral Realism: A Defence* (Oxford: Clarendon Press).

[62] Shorter, J.M. 1961. "Professor Prior on the Autonomy of Ethics." *Australasian Journal of Philosophy* 39: 286–287.

[63] Sidgwick, Henry 1907. *The Methods of Ethics*, 7th ed. (London: Macmillan).

[64] Sinnott-Armstrong, Walter 2000. "From 'Is' to 'Ought' in Moral Epistemology." *Argumentation* 14: 159–174.

[65] Star, Daniel 2011. "Two Levels of Moral Thinking." *Oxford Studies in Normative Ethics* 1: 75–96.

[66] Stocker, Michael 1971. "'Ought' and 'Can'." *Australasian Journal of Philosophy* 49: 303–316.

[67] ——— 1973. "The Schizophrenia of Modern Ethical Theories." *Journal of Philosophy* 73: 453–466.

[68] ——— 1979. "Desiring the Bad: An Essay in Moral Psychology." *Journal of Philosophy* 76: 738–753.

[69] Street, Sharon 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109–166.

[70] ——— 2011. "Evolution and the Normativity of Epistemic Reasons." *Canadian Journal of Philosophy* 35: 213–248.

[71] Tenenbaum, Sergio 2013. "Guise of the Good." *International Encyclopedia of Ethics* (Malden, MA: Blackwell).

[72] van Inwagen, Peter 1990. *Material Beings* (Oxford: Oxford University Press).

[73] Väyrynen, Pekka 2006. "Resisting the Buck-Passing Theory of Value." *Oxford Studies in Metaethics* 1: 295–324.

[74] Velleman, David J. 1992. "The Guise of the Good." *Noûs* 26: 3–26.

[75] Vranas, Peter 2010. "Comments on 'Barriers to Implication'." In *Hume, 'Is', and 'Ought'*, ed. Pigden (London: Palgrave Macmillan).

[76] Watson, Gary 1996. "Two Faces of Responsibility." *Philosophical Topics* 24: 227–248.

[77] Williams, Bernard 1973. *Utilitarianism: For and Against*, with J.J.C. Smart (Cambridge: Cambridge University Press).

[78] Wolf, Aaron *forthcoming*. "Giving Up Hume's Guillotine." *Australasian Journal of Philosophy.*

Vita

NAME OF AUTHOR: Aaron Wolf

PLACE OF BIRTH: Columbus, Ohio

DATE OF BIRTH: July 28, 1982

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Muhlenberg College, Allentown, Pennsylvania

DEGREES AWARDED:

B.A. in Philosophy, 2005, Muhlenberg College

PROFESSIONAL EXPERIENCE:

- Teaching Associate, Syracuse University
- Lecturer, Ithaca College