Virtuous Action, Inside and Out

by

Luke Gelinas

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Department of Philosophy

University of Toronto

© Luke Gelinas 2014
Abstract
This dissertation states and defends a novel view of virtuous action, motivated by a desire to do justice to the internal and external aspects of evaluating action. I argue that virtuous action is a complex notion, consisting in internal and external success. Acts are internally successful when they issue from admirable motivation, and externally successful when they achieve worthwhile or at least permissible ends. Virtuous action is well-motivated permissible success. I motivate the view by arguing that well-motivated permissible acts constitute organic wholes—entities whose value as wholes is greater than the value of their parts in isolation—and that the resultant view allows us to capture plausible ideas about the relation between virtue and benefit. I then defend the account from three rivals: (i) views on which subjective rightness supplies conditions on virtuous action; (ii) consequentialist theories on which virtue is a function of goodness produced; and (iii) views on which facts about virtue determine facts about rightness. I end by defending two claims about virtuous motivation. The first is variantism about the evaluative valence of individual motives: the fact that an individual motive is good (bad, neutral) in one context does not entail that it is good (bad, neutral) in other contexts. The second is that Aristotelian and Kantian accounts of virtuous motivation do not provide necessary, but may provide sufficient, conditions on all-things-considered good motivation.
For Rachael.

Sine qua non.
Acknowledgments

I arrived at the University of Toronto thinking I would write a dissertation on the medievals. A year later I was firmly hooked on contemporary ethics and the theory of virtue in particular. This was due almost exclusively to my encounter, in the very first class I took, with Tom Hurka and Sergio Tenenbaum, and later with Gopal Sreenivasan—the three who would come to form my dissertation committee.

I think of this collective as the committee *quo nihil maius cogitari potest*. I owe each a tremendous and indeed unpayable professional debt. Each one spilled much ink and many long hours guiding me through this project. I have learned a tremendous amount from them, and thoroughly enjoyed myself doing so. This project would not have been possible without them, and I am extremely grateful to each of them for the time and energy they invested in it.

I am also very grateful to Phil Clark and Mark Migotti, each of whom read large chunks of the project and provided valuable feedback at key points along the way; to Ingrid Stefanovic as well as Klaas Kraay, for their warm encouragement and support throughout my time at Toronto; and to John Hare, who first introduced me to ethics and provided a valuable model, early on, for what the philosophical life might look like.

On a more personal note, I am grateful to my mother, who continues to pour out love and support on a daily basis. And no less grateful to my father, for pushing me to think philosophically from an early age and doing everything in his power to open doors for his children; as well as my stepmother Valerie, for always believing in me. I must also mention my siblings, each of whom has influenced me for the better in their own distinctive ways; and my grandparents, whose company never gets old.

Throughout this project my three children—Sophie, Clara, and Asher—inspired me day after day with their creative energy and provided me with vast amounts of joy. You three are just about the brightest and coolest people I know. Though it may not be obvious on its face, this project bears your stamp as well.

Finally, I am most grateful to my partner Rachael, who has been a longsuffering source of encouragement and support. I couldn’t have even come close to finishing this project without you. Thanks for sticking with me.
# Table of Contents

**Introduction**  Yet another theory of virtue

0.0. Openings .......................................................... 1
0.1. Virtue and the good life ......................................... 2
0.2. ‘Virtue’ ............................................................. 6
0.3. ‘Right’ ............................................................. 7
0.4. Virtuous acts and virtuous traits .............................. 10
0.5. Overview .......................................................... 17

**Chapter one**  Virtuous action as an organic whole

1.1. Landscape ......................................................... 20
1.2. The Prichard-Ross approach .................................... 22
1.3. Is there substantive disagreement? .......................... 23
1.4. VA as an organic whole ........................................ 25
1.4.1. The very idea .................................................. 26
1.4.2. Motives ($m$) and permissible success ($s$) ............. 28
1.4.3. $w > m + s$ .................................................... 32
1.4.4. What sort of unity is needed? .............................. 37
1.4.5. Is permissible success necessary? ........................ 50
1.5. VA motivated ..................................................... 57

**Chapter two**  Further support

2.1. Virtue and subjective rightness ............................... 62
2.1.1. Sufficiency and subjective rightness ....................... 63
2.1.2. Is subjective rightness necessary? ......................... 65
2.2. Virtue and benefit 74
2.2.1. Personal well-being 75
2.2.2. Social goods 80
2.3. Cases 83

Chapter three  Virtue and consequence
3.1. Virtue-consequentialism 90
3.2. Stating virtue-consequentialism 92
3.2.1. Traits or motives? 92
3.2.2. Types or tokens? 93
3.2.3. What is it to have good (enough) consequences? 98
3.3. Virtue-consequentialism, right action 113
3.3.1. Goodness-based, rightness-based 114
3.3.2. Rightness and consequence 116
3.3.3. Non-exhaustion 119
3.3.4. Consequentialist rightness-based virtue 124
3.4. VA versus VC 128
3.5. Recap 132

Chapter four  Against virtue-based rightness
4.1. Entailment and explanation 133
4.2. Virtue-subjunctivism 134
4.3. Virtuous action-based approaches 141
4.3.1. Targeted theories 141
4.3.2. Agent-basing 145
4.4. P (rightness-based virtue) > P (virtue-based rightness) 152

Chapter five

Motive and context

5.1. Virtue and ‘admirable’ 155
5.2. A Rossian framework for individual good motives 158
5.3. Motive and context 162
5.3.1. Comparisons? 167
5.3.2. Variantism debunked? 171
5.4. Aristotle and Kant 176
5.4.1. Moralized Aristotle and Kant 177
5.4.2. Non-moralized Aristotle and Kant 179
5.5. Implications 188
5.5.1. Remainders and reasons 191
5.6. Conclusion 199

Works cited 201
Introduction

Yet another theory of virtue

From this tripartite distinction of the sects of philosophy, Marcus Varro, in his book De Philosophia, has drawn so large a variety of opinions that, by a subtle and minute analysis of distinctions, he numbers without difficulty as many as 288 sects—not that these have actually existed, but which are possible.

-St. Augustine, City of God, Book XIX

Since it cost a lot to win
And even more to lose
You and me better spend some time
Wondering what to choose.

-Hunter and Garcia, ‘Deal’

0.0. Openings

Augustine reports that Marcus Varro estimated the number of conceptually possible theories of the good life at 288. In Augustine’s day theories of the good life in effect doubled as theories of virtue, and vice-versa, owing to the widespread assumption that the concept of the ‘good life’ and the concept of ‘virtue’—arête or ‘excellence’—are intimately linked. Indeed it seems likely that the actual number of positions on virtue endorsed by philosophers in the history of thought cannot be far behind. Given this, it’s likely to strike many, as it often strikes me, as presumptuous at best, and sheer folly at worst, to affect to have anything worth saying about virtue.

Nonetheless, whether or not it constitutes something new, in what follows I develop and defend a theory about virtue. The view I develop is an account of virtuous action, one on which virtuous action is conceived as a whole with two properties or parts—good motivation and permissible
success—that combine to form a unified entity with its own distinctive value. Acts are virtuous, on this view, just in case and because they are both well-motivated and right.

\[\text{Virtuous Action (VA): Acts are virtuous just in case and because they are (i) well-motivated and (ii) right.}\]

Moreover, acts that satisfy VA constitute (so I argue) organic wholes: entities whose evaluative significance as wholes outstrips the evaluative significance of their parts in isolation.

The view I develop differs from many recent Aristotelian ‘virtue ethics,’ which haven’t lately been hard to find. Whereas many contemporary Aristotelians seem to think that virtue properties ground deontic properties, VA claims just the opposite: acts are virtuous partly because they are right.\(^1\) Here and throughout the sense of explanation in question is metaphysical. When I say that rightness ‘explains’ virtue, I do not mean merely that we can gain insight into virtue by understanding right action. While this is, I think, true, VA also maintains something stronger: that what makes acts virtuous is (in part) the fact that that they are right. This stands certain strands of contemporary ‘Aristotelian virtue ethics’ on their head (see chapter four).

A second way my approach differs from contemporary Aristeotelian views is its emphasis on virtuous action. Whereas many contemporary Aristotelians offer an account of what’s involved in having virtuous character, construed as stable virtues that persist over time, or elucidating the sort of good life or flourishing the virtuous enjoy, I focus on one-off virtuous actions.\(^2\) The reason I focus on virtuous actions rather than virtuous traits is because I think that virtuous action is at least as basic as virtuous traits, and that whether an act is virtuous does not depend on it issuing from a stable virtue (see below, sec. 0.3).

0.1. **Virtue and the good life**

At the same time, I share with many contemporary Aristotelians the conviction that being virtuous makes our lives go better, at least as measured on one scale. Indeed one motivation for the theory of virtuous action I defend is that it helps to explain an experience we commonly have when evaluating the goodness of lives.

---

\(^1\) The most prominent foil here is Hursthouse 1999.

\(^2\) And the foil here is Annas 1995.
Attempts to understand what’s involved in a good human life often combine an internal and an external element of evaluation. We often take the inner lives of persons (desires, motives, emotions, intentions, and so on) to be appropriate objects of praise and admiration, at least when they are calibrated in the right way. In general, what we’re like ‘on the inside’ matters. But we also assess human lives by more external criteria, such as whether we actually achieve our good aims; whether we in fact keep our obligations to others; whether our personal relationships and careers and projects come to fruition; and so on.

I think we can improve our understanding of the internal and external aspects of evaluating human lives on the assumption that virtuous action, as conceived by VA, comprises an essential part of a good life. It seems to me that part of the reason we evaluate human lives along both an internal and external axis, and the reason it’s appropriate to do so, is because it is appropriate to evaluate human action by both internal and external standards. On VA, internal success (acting from good motivation) and external success (doing what’s right) is the measure of virtuous or excellent human activity. On the assumption that virtuous action and the good life are linked, individuals who consistently engage in virtuous activity will, on this approach, be individuals who warrant positive judgments about how their lives are going on the internal and external axes—both of which, as I said earlier, we typically take to be relevant for living a good life. So I think VA can contribute to justifying our tendency to evaluate the goodness of human lives along both the internal and external dimensions.

The approach just sketched is also desirable insofar as it provides a nice way to bring some unity to the concept of ‘flourishing,’ as that notion is applied across human and non-human living things. Just as non-human living beings must actually achieve the proper ‘ends’ at which they ‘aim’ in order to flourish—as magnolias must actually bloom, and sea turtles must actually find their way to the proper beach and lay eggs, to be doing well as magnolias and sea turtles—so must humans actually achieve good aims to be good specimens of the human kind.

In humans no less than magnolias, moreover, luck and external factors can impede flourishing. A magnolia unlucky enough to receive too much sun and water (or not enough) will not achieve the ‘end’ of blooming, and so will not flourish. Likewise, a human who through bad luck does not realize the good ends at which she aims will not fully flourish. Welcoming the possibility that excellent human activity and flourishing requires co-operation from elements beyond our
control allows us to acknowledge that excellent human activity resembles excellent non-human activity in a basic way. This might in turn bring a greater degree of unity and coherence to the way we experience the world and our place in it.

The approach just sketched differs from Stoic views on the relation between virtue and flourishing, which deny that external factors can influence flourishing at all; as well as at least some interpretations of Aristotle, those on which whether we have the ‘external goods’ of fortune needed for flourishing is entirely distinct from whether we have virtue. Defenders of these approaches are likely to deny that the external aspect of evaluating lives and the concept of virtuous activity are linked, on the grounds that whether we do well in the external aspects often depends, unlike virtue and vice, on things beyond our control.

A big part of the motivation for these views is the (putative) connection between virtue (or vice) and responsibility, which would be broken if virtue and vice are luck-dependent. Breaking this connection seems to undermine the seemingly plausible view that virtue by its nature deserves praise, and vice by its nature deserves blame, since praise and blame are justified only for things within our control.

One way to respond is to deny that virtue and vice must be voluntary or within our control to be praise- and blameworthy. A second way is to concede that praiseworthiness requires responsibility but to deny that it is, first and foremost, praise rather than some other pro-attitude for which virtue calls—one whose appropriateness does not require responsibility, such as perhaps admiration.

Either of these routes seems to me more plausible than trying to insulate virtue from luck. Denying that virtue is luck-dependent flies in the face of what we know about the contingency of human goodness. Whether humans develop good character depends on many things beyond our control: genetic predisposition, emotional constitution, upbringing, social context, and the rest. These same factors can also hinder or prevent occurrent virtuous states, and virtuous motives in

---

3 Some interpretations of Aristotle are more hospitable to the thought that virtue is susceptible to luck—those, for example, that stress the passages in NE where Aristotle speaks of virtuous activity as actually hitting the target of virtue, rather than merely aiming at it. I take myself to be working within the general framework of human goodness and luck magisterially articulated by Nussbaum 2001.

particular, since whether our motives are good is often and in large part a function of psychological habits of thought and perception, how we are used to focusing our attention on the world, and (more mundanely) our moods—all of which are influenced by things we do not fully control.

While I have no knock-down argument for those who insist that it is always within our power to choose to act from a virtuous motive, we should at least request independent grounds for thinking so. From the perspective of the theory of virtue I think we do better to acknowledge the countless ways in which virtue rises or falls with happenstance—the ways it often depends on undeserved goodness, and is smothered by undeserved suffering—than to succumb to the urge to insulate human goodness from the vicissitudes of ‘step-motherly nature.’

For those who accept that, in general, human virtue is luck-dependent, the main motivation for denying that luck disrupts virtue by preventing the external realization of good aims and intentions is removed. There does not seem to be a relevant difference between luck disrupting virtue by hindering the formation of good motives and inner states, and luck disrupting virtue by preventing the realization of good intentions and aims. It seems to me that parity considerations of this general sort neutralize whatever motivation might exist for denying that luck undermines virtue by preventing the realization of good aims.

In addition to these assumptions about the relation between virtue and luck, my general framework for thinking about the issues that follow contains not always uncontroversial assumptions, many of which figure, more or less explicitly, in the arguments that follow. Two of the more important sets of assumptions concern the meaning and function of virtue and deontic terms. A third concerns the relation between virtuous acts and virtuous traits. I’ll say a bit about each.

0.2. ‘Virtue’

---

5 This not to say that virtue and vice fail to merit praise or blame, since (again) responsibility may not be strictly necessary for praise and blame. I remain neutral.
I take the English term ‘virtue’ to be synonymous with the Greek term *arête*, and to designate a certain type of value: the type of value that perfects something or makes it a particularly good instance of its kind. In what follows I use ‘virtue’ and ‘excellence’ interchangeably, to refer to the type of value that inheres in those things connected specifically with human activity or the will. That is, my topic in what follows is what has traditionally been known as moral virtue, rather than intellectual or physical excellence.

I find myself with moral realist inclinations, and tend to think that when we respond to virtue or excellence we are responding to a property or family of properties that are in some fairly robust sense mind-independent, and that justify and even demand certain responses from us, whether we recognize this or not. While I don’t want anything that follows to depend on realism about the value associated with virtue, I will assume that ascriptions of virtue imply claims about the appropriateness of certain reactive attitudes, most notably, the attitude of *admiration*. Whatever else virtue-judgments may or may not involve, they reveal a tendency to think well of something. I use ‘admiration’ to capture this thinking-well-of. It is the phenomenon we experience when considering the lives and deeds of Mother Theresa, Gandhi, Martin Luther King, Rosa Parks, Oscar Romero, Paul Farmer, Bill Gates, Jackie Robinson, Batman, and our favorite paragons of virtue generally.

To call *x* ‘virtuous’ is, I assume, to say that *x* is a particularly good or excellent or well-functioning member of its kind, in such a way that *x* warrants admiration. But what exactly is needed for something to count as an exemplary instance of its kind? I think it helps to distinguish two issues here. The first concerns the *degree of* value or goodness something must possess in order to render it a sufficiently excellent instance of its kind: just how good a motive or trait or person or life must be before it warrants virtue-ascriptions. In what follows I largely ignore this question; the explanation for why I do so is in chapter five (sec. 5.1).

The second issue concerns the different ways or aspects in which potential objects of evaluation can be good, and in particular whether something must be sufficiently good in *all* of its different aspects—good in all of the different types of ways it can be good—to count as virtuous. For example, traits of character can be good in more than one way. Traits can be good by disposing us to have good intentions, good motives, good desires, good emotions; to experience or ‘see’ the world a certain way; to experience appropriate reactive attitudes; to feel pleasure and pain in the
right way, at the right time, and so on. A trait that disposed its bearer toward good intentions but not toward good reactive attitudes would be a trait good in some respects, but not others.

More importantly for my purpose, it’s plausible that acts too can be good in more than one respect. Acts can fare well or poorly according to deontic standards, but they can also fare well or poorly according to the evaluative standards that govern the inner states from which they are done. While it’s controversial to what extent these two things can or should diverge, my view (which I argue for in chapter four) is that deontic-evaluation and motive-evaluation are distinct. But the general question is whether an object of virtue-evaluation, be it an act or trait (or person or life), must be good in all the respects in which it can be good to count as an exemplary instance of its kind.

Intuitively, it’s difficult to see how something could be a really good instance of its kind while faring poorly in some aspect. In what follows I assume that human action at least has two general aspects or ways in which it can be good, and that acts must be good enough in both of these ways to be virtuous. These two aspects are, as I said earlier, the deontic aspect and the motivational aspect. On my view, an act that fails to fare well in one of these regards will not be virtuous.

Those who find it compelling that objects of virtue-evaluation must fare well enough in each of their different aspects, and that acts can be evaluated along both a deontic and motivational scale, will have prior reason to accept the basic view of virtuous action I set out. This is indeed what led me to the view. However, while the good-in-all-aspects thesis is one of the considerations that in fact motivates the view, I have done my best to structure the arguments that follow in ways that do not assume its truth.

0.3. ‘Right’

VA claims that rightness explains virtue, but ‘rightness’ in what sense? Generally speaking, the more specific the notion of rightness employed in an explanation of virtuous action, the more ambitious, but also the less widely appealing, will be the resultant theory of virtuous action. If for example we specify rightness as the property had by all and only those acts that maximize

---

6 Here I am grateful to Gopal Sreenivasan for encouraging me early on to focus my attention on virtuous action, rather than virtuous traits.
utility, this will yield a much more ambitious and explanatorily powerful theory than if we specify rightness simply as the property that makes acts permissible.

I’ll proceed by employing the common distinction between giving an analysis of rightness at the metaethical level and offering a substantive normative theory of the right. I take metaethical theories of rightness to be concerned with analyzing the concept of rightness. They tell us whether rightness is a natural or non-natural property; whether ‘right’ denotes acts that are merely permissible or best; whether rightness is an objective or subjective notion; what the relation between the right and other moral notions is; and so on.

Normative theories of right, by contrast, supply us with substantive conclusions about the deontic status of particular acts; they tell us which acts are right, and what makes them right. Utilitarianism says that particular acts are right just in case and because they bring about a specified amount of utility; Kant tells us that particular acts are right when and because they pass the Categorical Imperative test; and so on.

In the initial phases of developing a rightness-based view, we should remain as neutral as possible on which substantive normative theory is correct, since the less encumbered the notion of rightness at issue, the wider the appeal of the theory. Once we have a convincing explanation of virtue in terms of general and abstract ‘metaethical’ claims about rightness, we can begin to argue over which substantive theory of rightness fits best with the specific claims about virtue we hope to capture. So I intend here to remain neutral on the normative features of the right and use relatively weak, metaethical claims about the right in the explication of VA. 7

The notion of rightness that, I claim, partly explains virtuous action has the following features. First, on the sense of ‘right’ at issue, acts are right just in case they are at least permissible. Right acts may be ‘better than’ permissible—may have stronger right-making features than acts that are merely permissible—but they need not be in order to count as right, in the sense relevant

7 It is thus important for my purposes, and I will take it as true, that it be possible to give a rightness-based theory of virtuous action without committing one’s self to fully sorting out the debate between different normative theories. This said, it’s neither possible nor desirable for rightness-based views to use a concept of right action that is entirely neutral between substantive lower-level theories, if only because the concept of rightness employed by rightness-based theories of virtue will of necessity be incompatible with the concept of rightness employed by virtue-based theories of rightness.
here. Second, the notion of rightness employed is the notion of all-things-considered, rather than merely *pro tanto*, rightness. In this sense VA is ambitious: it requires that virtuous acts actually be all-things-considered permissible, rather than merely having something to be said for them. Third, the relevant sense of rightness is objective: doing right acts, in the sense relevant to the theory I defend, turns on whether one’s actions are permissible in light of the facts.

I need to say a bit more up front about this notion of *objective* rightness. Objectively right acts are acts that the agent ought all-things-considered to perform, not necessarily from her own perspective, but rather in light of the facts. As I’ll understand it, an ‘objectively right’ act in this sense is an intentional action that aims to bring about an end that’s permissible in light of the facts, and that succeeds at realizing this end. This stands in contrast to theories of subjective rightness on which the agent’s perspective or the evidence available to her, rather than the perspective of the facts, determines what ought to be done. While the view I defend explicitly requires acts to be objectively right, it does not require them to be subjectively right. It’s worth pausing over this.

The most prominent accounts of subjective rightness make the subjectively right act at least partly a function of an agent’s beliefs. Just which beliefs are relevant, and whether anything more than beliefs matter, is controversial. One view says that the subjectively right act is the act best supported by all of the agent’s beliefs, both factual and moral. A second view says that the subjectively right act is the act best supported by the agent’s non-moral beliefs plus the correct moral principles. My view is that neither of these accounts identify a condition on virtuous action. These views do not identify a sufficient condition, since virtuous action requires good or admirable motivation; and it is possible to do the act one believes right, or the act best supported by all one’s beliefs, or the act best supported by one’s non-moral beliefs plus the correct moral principles, from bad motives as well as good.

While it’s more controversial, I argue later that these views do not identify a necessary condition on virtuous action, either. I do not, however, claim that *no* possible account of subjective rightness identifies a necessary condition on virtuous action. Unsurprisingly, it seems to me that the closer a particular notion of subjective rightness takes us to guaranteeing objective rightness, the more plausible it will be that it states a necessary condition on virtuous action. A limiting case of this would be an approach that defined the subjectively right act as the act best supported
by true beliefs about everything relevant to one’s choice—the empirical and moral facts—as well as true beliefs about what would come to pass were you to act this way rather than that. In this case your beliefs will always direct you to do the thing that actually turns out best, and subjective and objective rightness will be co-extensive. I discuss the relation between virtuous action and subjective rightness at length in chapter two.

0.4. *Virtuous acts and virtuous traits*

The last assumption requires the most discussion. (Indeed given that I am about to argue for it at length, it is not really an assumption at all.) In order for a theory of virtuous action to be worth defending, it must be true that virtue attaches to actions: actions must be genuine bearers of virtue. Moreover, the view of virtuous action I propose does not require acts to proceed from stable virtues, construed as excellent traits of character that persist over time, in order to be virtuous. Acts can be virtuous even if they do not proceed from, or stand in some other relevant relation to, stable virtues. This makes it necessary for me to defend two claims: (1) that acts are genuine bearers of virtue; and (2) that the virtue of acts does not depend on a connection to virtuous traits.

Are actions genuine bearers of virtue? While I haven’t been able to find examples of people who explicitly deny that acts are genuinely virtuous, some prominent views come close to implying this. Robert Adams, for example, claims that one fundamental difference between judgments of virtue and judgments of rightness is that the former are judgments about character, while the latter are judgments about actions.\(^8\) If judgments of virtue are always only judgments about character, and never judgments about actions, particular actions will never strictly speaking be virtuous. This conclusion is further suggested by Adams’ insistence that virtue is essentially a kind of *persistent* excellence in being for the good.

\[V\]irtue is not a merely momentary action or state of a person. One may be excellently, even heroically, for the good in a deed that takes but an instant, but that is not what we

mean by virtue. That is why the full formulation of my definition of virtue is that virtue is *persisting* excellence in being for the good.\(^9\)

If Adams is saying that actions by their nature are not the sort of things to exhibit persistent enough excellence to be the bearers of virtue properties, his view implies that acts are not, and cannot be, genuinely virtuous.

The difficulty is that common-sense indisputably recognizes acts as virtuous: we commonly speak of courageous acts, generous acts, just acts, and so on. Someone who denies that acts are bearers of virtue would be committed to thinking that such expressions (‘generous act,’ ‘just act,’ etc.) do not really ascribe virtue to the act, but are shorthand for something else. Perhaps the most plausible view of this sort is one on which calling acts generous or courageous is a shorthand way of saying that the act is an *effect* of a more basic bearer of the virtue-property in question. Calling something a ‘sunburn’ does not imply that the sun inheres in the burn, but rather that the burn is an effect of the sun; calling something a ‘generous gift’ does not imply that generosity inheres in the gift, but that it is an effect of a generous motive or trait.\(^{10}\) Likewise, calling something a ‘courageous act’ does not imply that the act itself is courageous, but rather that it is the effect of a courageous motive or trait.

This looks close to what many take to be Aristotle’s view, on which acts are virtuous only if they proceed from stable virtues. While it’s not clear that Aristotle actually held this view, given other things he says about acquiring virtues by performing virtuous acts, perhaps this is the best way to interpret Adams: as saying that virtuous acts inherit their virtue, and so must proceed from, virtuous traits. If so, Adams need not deny that acts can be genuine bearers of virtue. Someone who thinks that the virtue of acts depends on their being connected in the right way to virtuous traits can nonetheless maintain with full respectability that acts are genuine bearers of virtue. Acts will not, on this view, be *non-derivative* bearers of virtue, since their virtue will be at least partly explained by a connection to more fundamental virtue-bearers (i.e., traits). Nonetheless,

---


\(^{10}\) I owe these examples and this way of putting the view to Sergio Tenenbaum.
acts will still be genuine bearers of virtue: the mere fact that x’s being F can be explained in terms of y’s being F does not disqualify x as an F.

The real question, it seems, is whether the virtue of acts is derivative from the virtue of traits. There are different ways of understanding this idea. One way says that issuing from a stable virtue is necessary for an act to be virtuous: acts are virtuous only if they issue from virtuous traits. Views of this general type can disagree over whether issuing from a stable virtue is sufficient, but they all claim that it is necessary. A less ambitious view denies that acts are virtuous only if they proceed from stable virtues, while nonetheless affirming that an act’s issuing from a stable virtue makes it at least somewhat virtuous (or more virtuous). I’ll examine the more ambitious view first.

Actual arguments for why virtuous acts must proceed from stable virtues are, it turns out, hard to find. It seems to me that the only remotely viable route to the belief that virtuous acts must issue from virtuous traits rests on the idea that it is only actions that proceed from stable virtues that can have the properties needed for virtuous motivation.11 The story might go like this: to perform acts from the right motives—those necessary to qualify them as virtuous—the person acting must see the world aright, and respond appropriately. Virtuous motivation requires us to notice salient features of our surroundings and respond to them in the right way: with the right intentions, desires, motives, emotions, and so on. But the ability to see the world the right way and respond appropriately are acquired skills, had only by the virtuous. It isn’t possible for non-courageous people to perform genuinely courageous acts, since the sort of sensitivity and responsiveness needed for courageous motivation are had only by those with the skills of courage. Without these skills, genuinely courageous motivation is impossible. And once you have the skills, you have the virtue.

There is no need to deny that acquiring a virtue can resemble acquiring a skill, or that the way in which someone with a full-blown virtue and someone without it see and respond to the world can differ. It nonetheless seems clear that acts can be virtuous without proceeding from stable traits.

---

11 The following argument is inspired by things that Julia Annas 2011 and (to a lesser extent) Rosalind Hursthouse 1999 say. I do not know whether either would endorse it as stated here.
This point is recommended overwhelmingly by common sense.12 A soldier, concerned with the welfare of her comrades, throws herself on a grenade to spare them. Must we really think that this act is courageous only if the person is stably courageous? That seems incredible. No one would be inclined to withdraw a purple heart on finding out that this person did not consistently succeed at being courageous in other contexts.13 That’s simply irrelevant for whether she acts courageously here.

Or consider someone who, walking by a burning house, notices a toddler crying in a window. Suppose this person is generally timorous and averse to danger, even when important goods are at stake: the type of person who naturally shrinks from conflicts that he should, for the sake of justice, fight; refuses to climb ladders to rescue stranded animals from trees; and so on. None of this person’s friends would say he has the virtue of courage. Nonetheless, coming upon the house and noticing the toddler screaming in the window, smoke billowing from inside (with no one else in sight), the man finds himself overwhelmed with concern for the child’s safety. He also of course sees that he will be in great danger if he tries to rescue the child. In a rare surge of fortitude he runs into the blaze, locates the child, and carries her to safety.

This description of the case seems coherent. It also seems clear that, insofar as the man acts out of concern for the child, while appropriately assessing the danger to his person, he exhibits the motivation characteristic of courage. We can imagine the neighbors praising him for acting courageously, the local newspaper recognizing him for a courageous deed, and so on. These ascriptions seem perfectly in order.

Such examples seem to show that sometimes, at least, only minimal levels of motivational or practical reasoning skills are needed for virtuous motivation. It doesn’t take impressive powers of perception or fancy unified doxastic-orectic capacities to see that a toddler is in a burning house, realize that she needs help, and be motivated to save her. On the other hand, even if genuinely courageous motivation does require impressive sensitivity to morally salient features

12 Hurka 2006.
13 I owe this nice example to Tom Hurka.
of context, there seems little reason to deny that agents who lack stable virtues would, from time to time, and under the right circumstances, be capable of exhibiting them.

This brings us to a second argument for denying that virtuous acts must issue from stable virtues, one that provides a more direct response to the idea that virtuous motivation requires learned skills had only by the virtuous.

Assuming that proponents of trait-primary views accept a basic Aristotelian premise—that coming to be stably virtuous requires performing virtuous acts—coming to be courageous or gracious or just will typically involve performing acts characteristic of the trait. If performing these acts is to result in or contribute to possession of a virtue, the acts must, it seems, do more than merely outwardly resemble virtuous acts. Stable virtues guarantee reliability at performing virtue-appropriate acts by providing psychological materials that explain why the person with the virtue acts as she does; by supplying their possessors with the sort of motivation that is attuned to what matters, and so typically results in outwardly good acts. It would be a mystery how a learner could, by performing acts outwardly characteristic of a virtue, acquire stable dispositions to be motivated in the way needed to ensure reliably good action if the acts she performed during the learning process did not involve genuinely virtuous motivation. How could you come to be reliable at performing outwardly virtuous acts without being motivationally sensitive to important features of your surroundings? And how could you learn to reliably employ this sensitivity, to be consistently virtuously motivated, without ever actually being virtuously motivated?

Consider a parallel. It takes oodles of practice to make consistent contact with a baseball entering the strike zone at different trajectories and at speeds greater than 70mph. Learning to be a .300 hitter in this context—to make contact with the ball consistently, by the standards of baseball—will involve making solid contact with the ball on a number of occasions. But this will not be enough for you to learn to hit well. Someone who closes his eyes and whales away at every pitch might also make solid contact on a number of occasions, but will never, using this technique, become a reliable hitter.
Becoming reliable at the plate will, in addition to making solid contact on many occasions, also involve getting a feel for—in some sense grasping—what you did right on those occasions: a feel for how you experienced the situation and what you did that explains why you made contact. You will likely realize that keeping your back elbow up and the bat back increases your chances of making contact; that there are better and worse ways to set your feet; that the sooner you can visually track the ball out of the pitcher’s hand, the better. Once learners have a basic grasp on what works, they can begin to apply their technique and refine it in light of new experiences at the plate. The more advanced the learner, the more fine-grained the adjustments. Extremely good hitters will learn to gauge the trajectory of the ball by being able to see the way the seams on the ball spin as it approaches the plate; will learn to make last-minute adjustments with their forearms and wrists, to slap the ball to the opposite field; and so on. The point is that learning to be a good hitter requires emulating the right seeings, noticings, mental awareness, and judgments that result in hits. Since we cannot emulate the right internal states without experiencing the right inner states, learning to be a good hitter requires internal or broadly motivational success.

This plausibly resembles the type of process involved in acquiring a virtue. In the process of acquiring a virtue more than just repeated success at performing outwardly virtuous acts is needed. Just as a person will struggle to make consistent contact at the plate without ever seeing the ball right (and making the proper judgments about her stance, bat speed, and so on), so a person will struggle to consistently perform outwardly virtuous acts without ever seeing or being emotionally attuned to the situation in the ways needed for virtuous motivation. This suggests that the learner of virtue must, like the learner at the plate, actually get things right internally during the learning process. Someone who ‘learned’ consistently and reliably to perform outwardly virtuous acts without ever once getting it motivationally right would be like someone who ‘learned’ how to consistently hit the ball out of the infield with her eyes closed.

If this is right, the acts performed by learners will, in addition to being outwardly virtuous, also involve virtuous motivation. But if learning to be virtuous involves performing virtuously motivated acts, acquiring virtues involves virtuous motivation that is temporally and causally prior to the possession of the trait. Thus, virtuous motivation does not require virtuous traits, and it’s a mystery why acts must proceed from virtues in order to be virtuous.
Nothing I have said so far impugns the more modest version of the view that virtuous acts depend on virtuous traits. This second version concedes that acts can be virtuous even if they don’t express stable virtues, but maintains that the fact that an act expresses a virtuous trait imparts a measure of virtue to the act, making it at least somewhat more virtuous than it would otherwise be.

With one important qualification, this view is compatible with my own. The qualification amounts to specifying this idea in such a way that proceeding from a stable virtue is not sufficient for the act to be virtuous. If it were sufficient, VA would likely be false, since VA requires acts to be right in light of the facts. And it seems implausible that issuing from a stable virtue can guarantee this kind of external success.

At any rate, it seems to me that we have good independent reason to deny that issuing from a stable virtue suffices for a virtuous act, seeing as how this conflicts with the plausible empirical observation that virtues sometimes misfire in acts that aren’t virtuous. Even if virtuous traits dispose their bearers toward motivation that is often or usually good, and so toward virtuous acts, it seems too much to ask, given the moral complexity of the world, that virtues infallibly issue in good motives and acts.

A better statement of the view under consideration says that proceeding from a stable virtue can make virtuous acts—acts that already meet all of the relevant necessary conditions for virtue—more virtuous than they would otherwise be. While this view does not seem to me clearly true, it is not a battle I need to fight, since stated this way it is compatible with the view I defend. I’ll thus shelve the more modest version of the view and take the previous observations to warrant us to accept that proceeding from a stable virtue is neither necessary nor sufficient for virtuous action.

The claim that virtuous acts need not proceed from virtuous traits is compatible with a range of views on what role, if any, virtuous action plays in the explanation of virtuous traits, as well as on the nature of virtuous traits generally. I hope to remain as neutral on these questions as
Possible. I must leave the analysis of virtuous traits for another day; virtuous action is plenty to occupy me here.

0.5. *Overview*

The structure of the work looks like this. In chapter one I advance the basic claim about virtuous action. Above I stated VA as the claim that acts are virtuous just in case and because they are both admirably motivated and right. I precisify that as follows.

*Virtuous Action (VA):* Acts are virtuous just in case and because they are (i) all-things-considered admirably motivated and (ii) all-things-considered objectively permissible.

Because there will be natural doubts about what motivates VA, particularly in relation to views that make virtuous action a function solely of motive (what I call ‘the Prichard-Ross approach’), in chapter one I advance and defend an interpretation of VA according to which well-motivated objectively right acts constitute organic unities: wholes whose value is greater than the sum of their parts. In particular, I defend the claim that success at realizing good ends raises the value of well-motivated attempts, and/or that well-motivated attempts to realize good ends raise the value of success. This, I claim, gives us motivation to apply virtue-terms to acts that are both well-motivated and objectively right, since such acts will now have their own additional and distinctive form of value, one that merely well-motivated acts lack.

In chapter two I continue the defense of VA. I begin by arguing that objective rightness is a better candidate for explaining virtuous action than is subjective rightness, and indeed that subjective rightness fails to place any distinctive conditions on virtuous action. We can, I suggest, get everything that looks attractive about a subjective rightness condition on virtuous action from the condition that virtuous acts must be ATC well-motivated. I then argue that VA provides an attractive way of explaining two plausible claims about virtue: (i) that virtue is socially beneficial, increasing the goods of community; and (ii) that virtue is individually beneficial, increasing the good of their possessors. I end the chapter by considering cases. So far as the verdicts VA returns about virtue-judgments in particular cases go, VA is, I suggest, on a
par with the Prichard-Ross approach. I conclude that VA is a very strong competitor to Prichard-Ross style theories.

I take the arguments of the first two chapters to provide significant support for the conclusion that there is a tight logical connection between virtue and objective rightness. This does not, however, give us sufficient reason to accept VA, since there may be other views that both ensure a strong connection between virtue and rightness and that are more plausible than VA. One natural place to look for such a view, given VA’s emphasis on acts that succeed externally, is consequentialist theories of virtue that make virtue a function of goodness produced.

Chapter three is devoted to assessing these theories. It turns out that locating the best statement of virtue-consequentialism is surprisingly difficult. Statements of the view that make virtue depend on producing a net balance of intrinsic goods over intrinsic evils, and vice a function of producing a net balance of intrinsic evils over intrinsic goods, do not capture core consequentialist ideals: the relevance of preventing evil (to a state’s status as virtuous), and the relevance of preventing good (to a state’s status as vicious). The most natural alternative statement of the view claims that virtue and vice depend, not on the net production of intrinsic goods and evils, but rather on the value-difference states make to a world. This, however, raises its own, arguably worse, set of problems: it is very difficult, perhaps impossible, to move from plausible comparative judgments about the value of states to plausible judgments about their status as virtuous or vicious. I then go on to show that, even if we can state it adequately, virtue-consequentialism cannot ensure a necessary connection between virtue and objectively right action—even if act-consequentialism is the correct account of rightness. This is, I argue, problematic in its own right, and also suffices to show that virtue-consequentialisms cannot capture features of VA that make it attractive. While the net difficulties involved with virtue-consequentialism likely warrant us to reject it on its own terms, we are, I conclude, at the very least warranted to accept VA over competing virtue-consequentialist theories of virtuous action.

In chapter four I assess a class of direct competitors to VA, namely, views on which virtue explains rightness, rather than vice-versa. I consider what I take to be the three most prominent virtue-based theories of rightness currently in circulation: (i) virtue-subjunctivist approaches that
identify the right act as the one the virtuous person would do; (ii) targeted theories that make right action depend on ‘hitting the target’ of virtue; and (iii) agent-based theories that make rightness solely a function of motive. For each of these views a similar problem emerges. Their appeals to virtue are, I argue, otiose: in each case the facts on which putatively right-making virtue-properties depend are by themselves sufficient for rightness. This makes it simpler and more plausible to explain virtue by reference to rightness, or the features on which rightness depends, rather than vice versa. I conclude that, all things considered, we should prefer rightness-based virtue to virtue-based rightness. In conjunction with the arguments of chapters one and two, this gives us strong reason to accept VAs claim that the virtue of acts depends (partly) on their being right.

The fifth and final chapter is an explication of VA’s clause (i), the claim that acts are virtuous only if they are well-motivated. I say ‘explication’ rather than ‘defense,’ since I do not spend time arguing for this claim. It seems to me extremely difficult to deny that acts must be well-motivated to be virtuous. Instead I try to advance our understanding of the content of good motivation. Part of what emerges is an argument for contextualism or variantism about individual motives, understood as the view that an individual motive that is good in one context (such as the motive to keep a promise, or repair a wrong) may not be good in others. I then consider two historically prominent approaches to all-things-considered virtuous motivation: the Aristotelian slogan that virtuous motives consist in performing acts ‘for themselves,’ and the Kantian idea that virtuous motivation is to be identified with the motive of duty. I argue that no plausible interpretation of these views states a necessary condition on ATC good motivation, though they may state sufficient conditions. As an example of how developments in the theory of virtue can have implications for other important areas of debate, I end with some implications of variantism about individual motives for the theory of reasons and the relation between reasons and motives.

Without further ado, then, let me begin to get the actual arguments on the table.
Chapter one

Virtuous action as an organic whole

1.1. Landscape

In this chapter I set out a positive explanation of virtuous action. The view I advance claims that virtuous action is a complex notion, consisting in both motivational and deontic success. Acts are virtuous, on this view, just in case and because they meet two conditions, one concerning motives and one concerning rightness.

Virtuous action (VA): Acts are virtuous just in case and because they are (i) all-things-considered well-motivated and (ii) all-things-considered objectively permissible.

This chapter explains clause (ii); clause (i) is left for chapter five. While I accept VA, I will in what follows stop short of claiming that we have decisive reason to prefer it to other views of virtuous action—reasons that would or should compel all people to accept VA if they were seeing things right. My aim, here and throughout, is to make the case that VA provides an attractive and illuminating way to think about virtuous action, one that, while unappreciated, is capable of standing up to scrutiny and holding its own against other prominent approaches.

While the virtue-tradition as a whole has tended to accept a strong connection between virtue and right action, and while this connection is likely to strike many as intuitively plausible, there are at least two sources of resistance to clause (ii). The first stems from a common distinction with roots in the writings of H.A. Prichard and David Ross, according to which the goodness or badness of inner states, on the one hand, and the rightness or wrongness of acts, on the other, are entirely distinct. Acts can be well-motivated but wrong, and poorly motivated but right; the virtue of motives and the deontic status of acts do not depend on each other at all, in either direction. If we conjoin this with another claim distinctive of Prichard and Ross, that acting from good motives is sufficient for virtuous action, we land in conflict with VA. Since for Prichard and Ross the virtue of acts depends just on motive, and since motive and deontic status don’t depend on each other, virtuous action can be right, or it can be wrong—in which case rightness isn’t required for virtuous action and VA is false. Call this the P-R approach.
The second source of resistance comes from views that accept a strong connection between virtue and rightness but deny clause (ii)’s specification of ‘right’ in terms of all-things-considered objective permissibility. The feature of this specification likely to draw the closest scrutiny is the claim that acts must be *objectively* right in order to virtuous. For there is, many think, a subjective sense of rightness as well. As a first pass, objectively right acts are acts that are right from a bird’s or God’s eye perspective, rather than merely from the perspective of the person doing them. Subjectively right acts, by contrast, need not be right from a bird’s eye perspective (though they may be), but rather simply from the perspective of the person doing them: they are acts that should or may be done in light of our beliefs, credences, etc.

For example, whatever I may think or believe, if giving you the antidote in the green bottle will in fact kill you, and giving you the antidote in the red bottle will in fact save you, it is objectively wrong for me to give you the antidote in the green bottle, and objectively right to give you the antidote in the red one. By contrast, giving you the antidote in the red bottle need not be subjectively right, since I may have evidence both for thinking that the antidote in the red bottle will kill you, and that the antidote in the green bottle will save you. In this case it will be subjectively wrong for me to give you the antidote in the red bottle, even if it is objectively right to do so.¹

Given that we have a subjective as well as an objective conception of rightness, why think that performing virtuous acts requires doing what’s objectively right? Why isn’t doing the subjectively right act enough to qualify an act as virtuous? Subjective-rightness based theories of virtuous action agree with VA that virtue and rightness are closely connected, but insist that subjective rightness is the notion of rightness relevant for virtuous action. So the defense of VA runs through subjective rightness as well.

I’ll begin by considering the P-R approach. From there I motivate VA by offering an interpretation of it on which well-motivated objectively right acts constitute organic wholes with their own distinctive value. Most of the chapter is spent developing this idea and refining it in

¹ Defenders of objective rightness include classical utilitarians and their contemporary defenders, as well as certain deontological or rights-based theorists (perhaps most notably, Judith Thomson 1990: 170f). I will just assume that there is a distinction between subjective and objective rightness, and that which concept we employ depends on the sort of work we are trying to do.
the face of objections. Having done so, I then argue that, on certain important assumptions about the nature and functional role of virtue properties, these wholes are intuitively compelling candidates for virtue-terms, and for the concept of ‘virtuous action’ in particular. In the next chapter I tackle the relation between virtuous action and subjective rightness, and round off the positive motivation for the view by arguing that VA captures attractive ideas about the relation between virtuous action and benefit, while returning plausible judgments about particular cases.

1.2. The Prichard-Ross approach

The Prichard-Ross approach claims that the virtue of acts depends just on the inner states from which they are done. Taken in conjunction with another thesis that Prichard and Ross (and I; see chapter four) endorse, namely, that rightness does not depend on virtuous motivation, this implies that VA, and clause (ii) in particular, is false. If the virtue of acts depends just on motives, the virtue of acts does not depend on their being right. What reasons does P-R give for this claim?

The main reason Ross separated rightness from virtue was because he took virtue to inhere primarily in motives, and thought that the maxim expressed by ‘ought implies can’ precluded rightness depending on motives. Sometimes, perhaps often, we do not have it in our power to act from good motives—including times when the act in question is, intuitively, something we ought to do.

At best, however, the ‘ought implies can’ argument supports severing the dependence of rightness on virtue; it does not support severing the dependence of virtue on rightness. Ross is careless on this point. In The Right and the Good he says:

If it can be shown that nothing that ought to be done is ever morally good, it will be clear a fortiori that ‘morally good’ does not mean the same as ‘that ought to be done.’ Now it is, I think, quite clear that the only acts that are morally good are those that proceed from a good motive; this is maintained by those who I am now trying to convince, and I

---

2 The locus classicus for this view is perhaps Ross’s The Right and the Good (1930). I take Hurka 2001 and 2010 to provide a contemporary defense.
entirely agree. If, then, we can show that action from a good motive is never morally obligatory, we shall have established that what is morally good is never right.\(^3\)

This last sentence overreaches. If the ‘ought implies can’ argument works, it shows that it is never morally obligatory to act from virtuous motives. Defenders of VA can agree, by denying that doing what’s right depends on being virtuously motivated. It doesn’t follow, however, ‘that what is morally good is never right.’ For that we need a premise Ross has not given us: that the moral goodness or virtue of actions depends only on motive. What Ross actually says here—that ‘the only acts that are morally good are those that proceed from a good motive’—is consistent with maintaining that virtuous acts are also of necessity right, and so consistent with VA.

Does Ross give us some other reason to think that the virtue of acts depends just on motive? He doesn’t seem to. In the *Foundations* he says: ‘It is generally agreed among moralists that action owes its goodness, and the measure of its goodness, to the motive from which it springs.'\(^4\) If this is intended as the claim that the virtue of acts depends solely on goodness of motives, something more than an appeal to consensus or popularity is needed.

1.3. *Is there substantive disagreement?*

Are there other reasons to reject the thought that virtue depends on rightness? Proponents of P-R might simply claim that it is *a priori* or evident on its face that the rightness and virtue of acts are distinct, and/or *a priori* that acts are made virtuous just by virtuous motives. It may of course seem analytic or conceptually necessary to one person that \(p\), but not to another. Unsurprisingly the claim at issue—that the virtue of acts depends solely on motive—will not seem to defenders of VA anything like an analytic or conceptual necessity. I myself do not see how it follows from the concept of a virtuous act.

A more subtle P-R inspired objection to VA proceeds as follows. Both P-R and VA agree that good motivation and objective rightness are distinct: acts can be badly-motivated and right, or

\[^3\text{Ross 1930: 4.}\]
\[^4\text{Ross 1939: 313.}\]
well-motivated and wrong. On both P-R and VA, moreover, it will be best if one and the same act is both admirably motivated and right. The difference is that VA claims that admirable motivation and objective rightness are needed for virtuous action, while P-R says that admirable motivation is enough. But given that, for P-R, objective rightness still matters, this looks like a superficial disagreement. To see this suppose defenders of P-R stipulate that they will call acts that are both admirably motivated and right ‘overall superlative.’ P-R and VA will then agree that an act’s motives and its deontic status are components of overall act evaluation, and that the best or most excellent act is one that does well by each standard; they will merely disagree about what to call them. VA uses ‘virtuous’ to describe them, while P-R uses ‘overall superlative.’ In this light it’s not clear how much substantive disagreement even exists between P-R and VA, or what motivates VA in the first place.

A slightly different way to put the worry is that VA is a merely conjunctive view of virtuous action: it patches together two properties that everyone already agrees are operative in the overall evaluation of action—goodness of motive and rightness—and slaps the term ‘virtue’ on them. Nothing new is added by calling the conjunction virtuous; there are exactly the same evaluative factors present, in exactly the same degree, whether or not we call it ‘virtue’ or ‘virtue + objective rightness’ or ‘overall superlativeness,’ or whatever.

It seems right not to exaggerate cause for disagreement between VA and P-R. There is significant overlap between them, which we can welcome. Nonetheless, the implications of the two views differ. Perhaps most importantly, the views return substantively different conclusions about the relation of virtue to the right, and the relation of virtue to the good. Most obviously, P-R implies that the relation between virtuous action and rightness is relatively weak: on P-R doing virtuous acts never guarantees that one will do the right thing. VA posits a much stronger connection. Similarly, as I argue in chapter two, VA and P-R imply different conclusions about the relation between virtue and benefit, both individual and social. These differences matter. If the set of claims VA implies about the relation between virtue and the right and/or virtue and the good is more attractive than the set of claims implied by P-R, we have reason to prefer VA. The differences between VA and P-R also show up in the realm of individual cases, implying
different conclusions about the virtue-status of particular acts. Here we have reason to prefer the view that better captures our considered judgments.

Given that P-R and VA imply different things about the relation between virtue and other moral properties, the dispute between them is real. We should apply the concept of ‘virtuous action’ in ways that allow us to capture as many plausible hypotheses about the nature of virtuous action, including the relation between virtuous action and the right and the good, as possible. The way to decide between VA and P-R, then, is to trace out the implications of each view, and to see which set of implications is more plausible. This is one of my primary aims in what follows.

1.4. VA as an organic whole
While the dispute between VA and P-R seems significant, the charge that VA is a merely conjunctive view of virtuous action, and to that extent ad hoc, remains and requires a response. I think we can see that it is misplaced by considering the evaluative significance of the combination of good motives + objectively right action, and in particular by recognizing that this combination exhibits the sort of unity characteristic of organic wholes: entities whose evaluative significance as combinations is greater than the value of the sum of their parts.

This would accomplish several things. First, it would rebut the charge that VA is ad hoc. If the acts VA picks out constitute wholes of this sort, VA will not merely re-label two independent moral factors whose importance can be fully understood in isolation from each other. VA will still be a conjunctive theory of virtuous action, but now the conjunction it picks out will be significant in itself, as a combination. Second, it would distinguish VA from P-R, since P-R does not capture the significance of these wholes beneath the concept of virtue. Third, it would offer a more direct source of support for applying ‘virtuous action’ to well-motivated objectively right acts, since these acts will now be loci of extra value—whatever value they have as wholes. Given that virtuous acts are particularly good or excellent acts, the extra value residing in these acts will bolster their claim to the mantle of ‘virtuous.’ The organic unity interpretation will also, more broadly, give us a new and productive way to think about the combination of admirable motives and objective rightness, even for those who deny that both are involved in virtuous action.
There are two independent sources of support for the claim that admirable motivation + objective 
rightness constitutes an organic whole. The first is case-based: our judgments in particular cases 
are best explained on the assumption that these wholes are present. The second is more general, 
appealing to the idea that there is the right sort of fit between well-motivated attempts to achieve 
permissible ends and actually achieving permissible ends: the type of fit characteristic of 
paradigmatic cases of organic unity. I’ll motivate the view with some cases before turning to the 
more general considerations. But before that I need to deal with an objection to the very idea 
that facts about rightness and facts about intrinsically good motives could combine to form 
organic wholes.

1.4.1. The very idea

It might seem difficult or impossible for VA to pick out an organic whole, since the notion of 
organic wholes, as it comes to us from Moore, is a claim about intrinsic value. An organic whole 
is an entity whose intrinsic value—not instrumental or some other form of value—outstrips the 
sum of the value of its parts. On the assumption that a part could not contribute to the intrinsic 
value of a whole by being anything other than intrinsically valuable, this implies that the parts 
contribute to the value of the whole by themselves being intrinsically valuable. If so, the claim 
that admirable motives + permissible success constitute an organic unity entails that both 
admirable motives and objective rightness are intrinsically good. But while nearly everyone 
agrees that motives can be intrinsically good, the claim that right acts are intrinsically good is 
more controversial. Certain deontologists deny it, and others are likely to as well. 5

The organic unity interpretation does not, however, require objectively right acts to be 
intrinsically good. A number of writers have pointed out that the basic idea behind organic 
unities, that the moral importance of a combination is sometimes not reducible to the sum of the 
importance of its parts in isolation, applies to moral factors more generally. 6 Moral factors and 
distinctions relevant for determining rightness, and which need not be reducible to intrinsic 
goodness, can vary in force, depending on their combination with other factors.

For example, many want to recognize cases where doing harm is worse than allowing harm. This tempts us to think that doing harm is always worse than allowing harm. Shelly Kagan points out that, in cases of self-defense, this doesn’t seem true. If you charge at me with a knife intent on killing me, it is arguably no worse for me to push you into a pit, thereby harming you, than it is for me to let you fall unwarned into the covered pit in front of me, thereby allowing you to be harmed. Kagan claims that, if we do agree that doing harm in this case is no worse than allowing harm, the best explanation is that the moral force of the do/allow distinction depends on whether we are in a context involving self-defense.

Along similar lines, Frances Kamm argues that the weight of certain factors associated with killing and letting die varies, depending on what other factors are present. Kamm suggests that the negative significance of letting someone die depends on how much effort or personal sacrifice it would require to save him, such that the easier it is to save someone from death, the worse it is to let her die. Here the factor of personal cost interacts with, and changes the force of, the factor of letting die.

On these views the combination of moral factors per se, rather than just the factors considered on their own, is significant. Moreover, these factors need not be understood in terms of intrinsic value for this to be true. If they are understood in terms of intrinsic value, they will be Moorean organic wholes. But if they are not, we will still have combinations of morally significant factors whose important or force is not reducible to the sum of their significance in isolation. If Moore’s original insight into organic wholes is really an insight into the basic workings of moral properties generally, not just intrinsic value, those who deny that objective rightness is intrinsically valuable can still accept the organic unity approach to VA.

---

7 Kagan 1998: 18. Kagan calls the assumption that moral factors always have the same weight the Additive Fallacy.
9 Kamm 1996: 56.
10 We could also note the moral particularism defended by Jonathan Dancy and others, which similarly holds that moral factors behave differently in different contexts, depending on what other considerations are present. Dancy 2004; Hooker and Little 2001.
11 CF. Kagan: “Indeed, my entire discussion of the additive assumption can be viewed, with hindsight, as a generalization of Moore’s point to all of ethics, together with an application of this generalization to areas where its relevance has not been recognized” (1998: 23).
Some might, however, find this too quick. It isn’t immediately obvious that the examples from Kagan and Kamm show what I need them to show, namely, that mixed organic wholes, combining intrinsic values with deontic properties, are possible. For it might be thought that the combinations in the examples from Kagan and Kamm—the significance of killing/letting die in contexts of self-defense; the badness of letting die and the costs of saving—do not involve intrinsic values at all, but deontic factors irreducible to intrinsic values. In that case we would have organic wholes consisting in combinations of two deontic factors, not (as the mixed interpretation of VA claims) one deontic factor plus one intrinsic value.

While it is, I agree, possible to understand the relevant deontic factors as irreducible to intrinsic values, the force of the examples does not seem to depend on it. Suppose that in Kamm’s example we understand the ‘sacrifice’ or costs involved in saving someone from death in terms of the degree of pain or other intrinsic disvalue suffered in the life of the rescuer, while denying that the negative significance of letting another die can be fully captured by appeal to the intrinsic badness of the death. In that case letting die is a deontic factor irreducible to intrinsic disvalue. Such a combination of views seems perfectly possible. But someone who held this while at the same time denying the possibility of mixed organic wholes would be forced to deny Kamm’s insight that the negative significance of letting die can depend on the cost of saving. This seems unmotivated: we do not lose the intuition that the negative significance of letting die can depend on the degree of sacrifice it requires merely by discovering that personal sacrifice is to be understood in terms of intrinsic disvalue, while letting die is not. The intuitive force of the examples from Kagan and Kamm do not depend on how we understand the nature of the factors involved. Someone who insists that mixed views of organic wholes are not possible must explain why we should ignore the continuing intuitive force of the examples when they are cashed out this way.

1.4.2. *Motives (m) and permissible success (s)*

The organic whole approach to VA can, I conclude, take one of two forms. It could, first, insist that both objective rightness and admirable motivation are intrinsically valuable, and claim that the whole composed of admirable motives + objective rightness is an organic whole in the

---

12 I am grateful to Tom Hurka for this objection.
original, Moorean sense. Or it could simply remain agnostic on whether objective rightness is intrinsically valuable, advancing instead the less determinate claim that the evaluative significance (not necessarily intrinsic value) of admirable motives + objective rightness is not exhausted by the significance of these things in isolation. While my view is that the importance of objectively right acts consists at least partly in their ability to make our lives and the lives of those around us better, and is profitably conceptualized at least partly in value-terms, someone who disagrees can still accept the organic unity interpretation of VA.\(^\text{13}\)

I will take it for granted that motives are capable of being intrinsically good or intrinsically bad, and that the value of the motive/s from which an act is done make a difference to the moral significance of an act. Acting or attempting to realize an end from intrinsically good motives is so far forth good, while attempting to do so from intrinsically bad motives is so far forth bad. Call the value that attaches to attempts by virtue of the goodness or badness of the motives from which they are done ‘motive-value,’ or \(m\).

The claim that realizing permissible ends is significant or valuable is more complex. It is, for one, distinct from the idea that the importance of objectively right action is exhausted by the value of the good states of affairs in which it often (but not necessarily always) results. For the success itself can also be valuable, in addition to the value of any states of affairs realized.

The value of this success might depend on different things. It might be partly a function of the extent to which realizing an end is complex and/or difficult;\(^\text{14}\) or the extent to which the agent has a justified belief that an end is within reach, so that the success is non-lucky;\(^\text{15}\) or the extent to which it involves the redemption of prior self-sacrifices;\(^\text{16}\) or whether the end realized is good or worth realizing. These views can also be combined, so that whether or the extent to which a success is valuable depends on non-luckiness and complexity and the redemption of prior sacrifices, and so on.

\(^{13}\) Though they may not be able to accept all of the conclusions of sections 1.6, below.
\(^{14}\) Hurka 1996a and 2006; Bradford 2013.
\(^{15}\) Hurka 1996a.
\(^{16}\) Portmore 2007.
These constraints on the value of success are often found in theories of achievement. For our purposes, however, it will help to keep discussions of achievement distinct from discussions of successfully realizing aims. Not all successful realizations of ends need count as achievements, but this on its own doesn’t imply that the former aren’t valuable. Since objectively right action can be non-complex, non-difficult, and involve the redemption of no prior self-sacrifice, it will be best to consider simple cases of success, as this will help block out irrelevant noise.

One question that I cannot gloss over is whether success at realizing ends is valuable only when or to the extent that the ends realized are good or worth realizing, or whether realizing any old end is valuable.\textsuperscript{17} If success at realizing any old end is valuable, then performing objectively right acts can be valuable without it being facts about rightness that explains this. The importance might be fully explained by the simpler fact that objectively right action is the realization of some end or other, rather than an end that’s objectively right. In that case it would be success at realizing ends that enters into the organic whole, rather than success at realizing right ends.

While it is compatible with the organic unity interpretation of VA to think that realizing any old end is significant—since the importance of objectively right action might be a function of two facts: that the act realizes an end, and that the act realizes an end it’s permissible to realize—it seems to me implausible that realizing any old end is per se valuable. This despite recent defense from Simon Keller, who appeals to the intuition that it is preferable to achieve meaningless goals (such as counting the blades of grass on one’s lawn), or indeed even downright evil ones (such as torturing puppies for pleasure), than trying but failing to do so.\textsuperscript{18}

I agree that achieving certain inherently meaningless ends, such as crossing a line 26.2 miles away or standing atop a mountain, can be valuable.\textsuperscript{19} But this is, I think, only insofar as these ends are permissible. It seems to me much more problematic to claim that realizing wrong ends

\textsuperscript{17} Rawls 1971, ch. 7; Raz 1986, ch. 12; Adams 1999, ch. 3; Wolf 1997; Scanlon 1998, ch. 3; Crisp 2000. Here it’s important to make a distinction between realizing ends that are intrinsically good (making another happy, etc.) and realizing ends that it is good or permissible to realize. The latter does not amount to the former, since sometimes we are permitted and indeed obligated to realize states of the world that are intrinsically bad, such as when causing someone pain is necessary to prevent even greater pain.

\textsuperscript{18} Keller 2004.

\textsuperscript{19} Hurka 2006.
is valuable. If realizing wrong ends is valuable, there will always be something to be said for wrong acts. We will *always* have reason to act wrongly: whatever reason is generated by the value of realizing ends generally. But it certainly seems possible to have no reason to do the wrong thing. The view that there is always some reason to act wrongly requires a controversial metaethical claim about practical reasoning: that reasons against performing the right act are always outweighed, in such a way that their force persists, rather than being canceled or ‘silenced’ by right-making considerations.

More importantly, even if we grant that a hydraulic or weighing conception of reasons is true, when wrong acts *do* have something to be said for them, this does not in any way seem to be explained by the prospect of their success. The possibility that we might succeed at making the dog suffer, or ridiculing a child to tears, or assassinating a well-loved politician, does not weigh in favor of doing so. I will have more to say below about the relation of impermissible success to the wholes I have in mind.

While I cannot here offer an extended theoretical grounding for our sense that permissible success is positively valuable, I will assume that the thought enjoys enough intuitive support to warrant its role in the argument that follows.²⁰ Call the value that attaches to simple cases of objectively right action by virtue of realizing good or acceptable ends ‘success-value,’ or $s$. We can now say that the value of any particular well-motivated objectively right act will be no lower than the value that accrues to the act by virtue of its success at realizing good aims ($s$) plus the value that accrues to the act by its being admirably motivated ($m$). If we let ‘$w$’ designate the value that accrues to an act by its being a whole consisting in admirable motivation plus permissible success, we can then say that the question of whether admirably motivated objectively right action constitutes an organic unity is the question of whether $w > m + s$. The

---

²⁰ The theoretical grounding I prefer is similar to one advanced by Keller, who argues that certain attitudes contain within them their own standards for success (Keller 2009). On Keller’s view the attitudes involved in having a goal are by their nature such that they succeed when the goal is attained, and frustrated when not. Since it is better if our attitudes succeed than fail, it is better when we realize goals than when we do not. I differ from Keller in holding that the constitutive aim of the attitudes that comprise goal-seeking is not realizing aims *per se*, but realizing *good* ends. This gives me a route to the conclusion that realizing good or permissible ends is valuable, while realizing bad ends is not—though more argument would be needed.
organic whole interpretation of VA claims that, sometimes at least, \( w > m + s \), while those who deny it claim that \( w = m + s \).

1.4.3. \( w > m + s \)

Is it plausible that \( m + s \) constitutes an organic unity? There are two sources of support for this idea: one more abstract and one more concrete. The more abstract support comes from reflecting on the relation between well-motivated attempts and objective rightness, allowing us to see that \( m + s \) has the right sort of fit to constitute an organic whole. The more concrete source of support comes from reflecting on particular cases where we cannot explain everything we want to without appeal to the idea that \( w > m + s \). I’ll begin with the cases; the more general considerations emerge below.

There are two ways \( w \) could be greater than \( m + s \), corresponding to two different views of the nature of organic unities.\(^{21}\) On the first view, Moore’s own, the extra value resides in a whole, \( w \), constituted by the parts, \( m \) and \( s \). On the second approach, often called the ‘conditional’ view, the presence of one feature, \( m \), boosts the value or significance of the other, \( s \). Here the extra value is located in the original factors themselves, rather than in a whole constituted by them. I do not think anything that follows turns on the choice between the Moorean and conditional approaches, and will therefore remain neutral. While I will frequently write as if I assume the conditional view, this is just for ease of expression; everything I say in what follows can be reframed in terms of the Moorean view.

I will proceed by *reductio*. Suppose it is false that \( w > m + s \). Two things follow: (1) the value of a well-motivated attempt to realize a permissible end depends in no way on whether the attempt succeeds, in which case the value of a well-motivated attempt that succeeds will be the same as the value of an identically well-motivated attempt that fails; and (2) the value of permissible success depends in no way on whether it is well-motivated, in which case well-motivated success at realizing an end will be no more valuable than badly-motivated success at realizing the same end. If either (1) or (2) is false, we can conclude that \( w > m + s \).

\(^{21}\) Hurka 1998.
Take first (2), the claim that the value of realizing a permissible end depends in no way on the motive involved. Suppose you accomplish some good end: you make your partner happy by preparing her favorite meal. Setting aside the value involved in the result of your act (the actual happiness of your partner), what is the value of your accomplishing this goal, of making your partner happy?

It plausibly depends on why you are trying to make her happy. Imagine you do so solely out of self-interest: you know that if your partner is happy, she is more likely to vacuum than when she is in a bad mood. You don’t take any pleasure in your partner’s happiness, or want her to be happy for itself; you just don’t feel like vacuuming. Or, worse, suppose you try to make your partner happy because you know that when she is happy she is easily deceived, thus making it easier for you to sneak off with another love interest.

We can all agree that the presence of bad motives makes acts so far forth bad. The question is whether the presence of bad motives influences the value that accrues to the act by virtue of it being the realization of a permissible end. Some may be tempted to conclude that there is no positive value, or at most very little, in the success—making your partner happy—when your motives are bad. But strictly speaking I don’t need to claim this. To show that motives matter to the value of success, it just needs to be true that well-motivated success is more valuable than badly-motivated success, which is compatible with badly-motivated success having some value.

It seems clear that well-motivated success is more valuable than poorly-motivated success. Suppose again that you make your partner happy, but that now you are motivated by love and a desire for her happiness. Here making your partner happy seems to have significant positive value—more value than making her happy because you don’t feel like vacuuming, or because you want to sneak off with another love interest. If making your partner happy because you love her is better than making her happy from bad motives, (2) is false: the value of success depends on the value of the motives behind it.

We might be tempted to resist this conclusion on the grounds that we can explain the difference between well-motivated and badly-motivated success simply by appeal to the difference in the

---

22 It might be objected that the end of making one’s partner happy in these examples is not permissible, precisely because your reasons for doing so are bad. Here I simply assume that it is possible to do objectively right acts from bad motives; see chapter four for argument.
value of the motives. But this doesn’t capture our sense that the value of the success itself varies, depending on motive. It is the realization of the end itself, the making our partner and friends happy in its own right, that seems more valuable when our motives are good than when they are not.

A second reason not to explain the difference between well-motivated and poorly-motivated success just in terms of motive-value involves us in consideration of (1): the thought that the value of well-motivated attempts does not depend at all on whether they succeed. If well-motivated success is better than badly-motivated success simply because the former is better-motivated than the latter, as the objector claims, well-motivated failed attempts to realize permissible ends should be better than badly-motivated failed attempts, simply by virtue of the difference in motive-value. That is, the value of attempts to realize permissible ends should depend just on goodness of motive.\textsuperscript{23}

But if the value of attempts to realize good ends depends just on motive, well-motivated attempts that fail to realize good ends will be just as valuable as well-motivated attempts that realize good ends. This seems false. Consider a case where you are motivated by love to prepare your partner’s favorite dish, but your attempt fails. The key ingredient in the dish you are preparing has gone bad on the shelf, and the meal is a failure. If well-motivated success at making others happy is not per se valuable, it will be just as good for you to be well-motivated and fail to please your partner as to be well-motivated and actually please her. But actually making your partner happy because you love her seems valuable. It is better if our well-motivated attempts to express love and do good end in success than in failure. We are invested in our attempts and tryings in such a way that we prefer them to come to fruition, and have at least some slight occasion for regret or frustration when they do not. In that case (1), like (2), seems false.

One natural objection at this point is that the rationality of preferring well-motivated success to well-motivated failure can and ought to be explained simply in terms of the valuable states of affairs that permissible successes do, and failures do not, realize. The reason we prefer a well-motivated success is because of the good it brings about, which isn’t present in a well-motivated failure.

\textsuperscript{23}If it did not, explaining the difference between well-motivated and badly-motivated success solely in terms of the difference in motive-value would be a non-sequitur.
It is, however, controversial whether all right acts make the world better. Some hold that the duty to keep one’s promises, for example, is not reducible to a more general duty to promote the good. If so, keeping promises can be right even if it does not make the world better. But well-motivated success in these cases still seems better than well-motivated failure. Actually keeping a promise because you respect your word and the person to whom you give it seems better than trying but failing to keep a promise from the same motives, even if keeping the promise does not leave the world better-off.

Moreover, even in cases where objectively right acts do bring about valuable states of the world, we cannot capture everything we want without appealing to the importance of success itself. Imagine that your partner turns out just as happy in the scenario where you fail to make her favorite dish as in the scenario where you succeed. Unbeknownst to you, a friend has decided to surprise you both by having a local restaurant deliver dinner, which contains a splendid version of your partner’s favorite dish—the one you just failed to make. Your partner takes as much pleasure and happiness in the restaurant’s version of the dish as she would have in yours. While you have not accomplished your aim of making your partner happy, she is just as happy as you hoped she would be.

Nonetheless, there seems something less desirable about this scenario than one where you succeed at making your partner happy. It is undesirable that your well-motivated attempt to do something nice for her has failed.

This isn’t necessarily to say that we should always be less pleased when those we love are made happy by things we aren’t responsible for, or that it is best for us to be motivated by the thought that making the dish is a way for me to make my partner happy, rather than simply that she will be happy. To insist that there is special significance in my making the people I love happy might appear self-centered or egocentric. What should matter to us is just that our partner be happy; how this comes about should make no difference.\(^\text{24}\)

But even if it is better to be motivated purely by my partner’s happiness than by my making her happy, the latter motive may still be significantly good. Indeed the desire to please those we love seems distinctive and even constitutive of loving relationships. Those in loving

\(^{24}\) I am grateful to Sergio Tenenbaum for this objection.
relationships often share a history that makes them particularly well-suited to benefit each other, and to do so in ways that are attentive to subtleties of the other’s personality and past. This can imbue attempts to please loved ones with expressive significance and value that is not present in cases where luck or the world succeed at making them happy, and is part of what constitutes the relation between lovers—part of what makes you and her an *us*. It would be a shame, even tragic, if we never succeeded at pleasing those we love, even if they were not any less happy as a result, and even if we concede that the ideal motive does not involve reference to ourselves.

While the preceding argument has been somewhat complex, the basic phenomenon I have in mind can be observed in familiar and more striking cases of well-motivated objectively right action. Consider Nelson Mandela’s efforts to end apartheid in South Africa. It seems plausible that through his efforts Mandela was partly responsible for ending apartheid, that his attempts to do so were objectively right, and that he was well-motivated in these attempts—motivated by the unjustness of apartheid and the suffering of those subject to it. It seems difficult to explain the overall value of Mandela’s successful efforts in terms of these factors in isolation. The fact that Mandela’s well-motivated efforts ended in success makes a positive difference to the way we view those efforts. If they had failed Mandela’s attempts, while admirable, would not be esteemed in the same way. We would be more likely to consider his life a waste, perhaps tragically, and to think that there would have been better ways for him to have spent it. The fact that Mandela’s admirable efforts succeeded make the efforts seem even more valuable. Likewise, the fact that Mandela’s attempts were well-motivated seems to matter to the value of his success. If Mandela had been motivated solely by fame or greed, we would not regard his accomplishment as highly as we in fact do.\(^{25}\)

While the case of Mandela involves a number of confounding variables, such as the complexity and difficulty of attaining the end, and the amount of personal sacrifice required of him, it does illustrate the basic plausibility of the claim I’m defending. Having elicited the judgment that well-motivated achievements like Mandela’s are valuable, we could, if we desired, begin to factor out the confounders, focusing on whether we would reach similar conclusions about the value of Mandela’s efforts had they not been complex, or involved personal sacrifice. I think we will find that in such simpler cases realizing good aims is still valuable (though perhaps not as

\(^{25}\) I am grateful to Tom Hurka for this example.
valuable), and that its importance cannot be fully understood apart from facts about motive. It is an attractive feature of VA that it shows us how our sense that impressive achievements like Mandela’s are valuable can apply more widely to everyday accomplishments in all spheres of life.

1.4.4. **What sort of unity is needed?**

I said earlier that, in addition to particular cases, there is a more abstract source of support for the idea that good motives + permissible success forms an organic whole. If at a suitable level of abstraction the relation between good motivation and permissible success is of the right sort to comprise an organic whole, we will have a further source of support for the claim that $w > m + s$, one that can help to ground and bolster our intuitions about the preceding cases.

What would count as evidence for the claim that the relation between good motives and permissible success is of the right type, the sort that characteristically obtains between elements of organic wholes? One way to proceed would be to advance a general account of this relation, and then to see whether the relation between $m$ and $s$ is an instance of it. Alas, I have no general account of the relation between elements of organic unities. Indeed I am not sure there is a unified account of this relation; the relation might differ, depending on the nature of the entities involved. I’ll proceed instead with an argument from analogy, comparing the relation between good motivation and objective rightness with the relation that characterizes what many take to be a paradigmatic case of organic unity: the enjoyment or appreciation of genuine beauty. I will simply assume that the appreciation of genuine beauty does form an organic whole, and that resemblance to the relation that characterizes this whole—the relation between aesthetic enjoyment and genuine beauty—gives us evidence for thinking that a relation is of the right sort.

Assuming that the appreciation of true beauty forms an organic unity, it does so by combining an intentional attitude, enjoyment or appreciation, with an intentional object, genuine beauty. What is the nature of the relation between this attitude and this intentional object? One answer is that genuinely beautiful things are something like the ‘formal object’ or *constitutive aim* of aesthetic appreciation, in much the same way that (the story goes) the true is the constitutive aim of belief,
or the good the constitutive aim of practical reasoning. What exactly we mean by saying that $x$ is the constitutive aim of some attitude or activity, not to mention claims about what is the constitutive aim of what, are likely to be controversial. I do not want to rest too much on them. But there are at least a couple ideas involved in the concept of a ‘constitutive aim’ that might help us think more clearly about the relation between the elements of organic unities, even if we ultimately deny that this relation is best captured in terms of it.

First, to call true beauty the constitutive aim of aesthetic enjoyment is at least to say that the beautiful is a proper object of enjoyment. Aesthetic enjoyment of the genuinely beautiful is warranted or fitting, in a way that aesthetic enjoyment of the ugly is not.

Second, and closely related to the first point, is the idea that facts about the constitutive aims of attitudes and activities have a role to play in determining what sort of considerations give one the right sort of reasons for having the attitude or engaging in the activity. It at least seems true that, if beauty is the constitutive aim of aesthetic enjoyment, that an object is beautiful gives us grounds for aesthetically appreciating it. Some are also likely to accept the stronger claim that, if beauty is the constitutive aim of aesthetic appreciation, the only appropriate grounds we could have for aesthetically appreciating an object is that it or certain of its features are beautiful. Considerations not tied to the beauty of objects would, on this view, not provide the right kind of grounds for aesthetic appreciation, much the same as (it is sometimes said) non-evidential reasons for belief—such as the fact that believing in God will have practical pay-off—are not the right grounds for belief. This latter claim is, however, controversial. Some people deny that constitutive aims impose limits on reasons in this way. I will not rest anything on this stronger claim in what follows, but instead focus on the weaker sufficiency claim: if true beauty is the constitutive aim of aesthetic enjoyment, then that an object is beautiful gives us grounds for appreciating it.

---

26 I don’t mean to endorse these claims about the true and the good; just to illustrate what is supposed to be involved in calling something a ‘constitutive aim.’
27 I assume that there can be different types of enjoyment or appreciation, and that the truly beautiful is an appropriate object of some, but not necessarily all, of them. I have no further account of what distinguishes different types of appreciation.
28 See the debate initiated by Rabinowicz and Rønnow-Rasmussen 2004.
Third, the constitutive aims of attitudes or activities tell us not only what aims it is appropriate for these attitudes or activities to have—that it is appropriate for belief to aim at the true; practical reasoning at the good; and aesthetic appreciation at the truly beautiful—but also imposes limitations on what can count as instances of the attitude or activity in question. The exact way in which constitutive aims impose limitations is (again) likely to be controversial, and might vary by case. With respect to enjoying the beautiful I think we can at least say that, if the beautiful is the constitutive aim of aesthetic appreciation, aesthetic appreciation of an object that one knows is not beautiful is incoherent. Someone does not count as engaged in the activity of aesthetic appreciation unless it is genuine beauty she aims to admire. And someone who aims to admire genuine beauty cannot admire objects (or features of objects) she knows are not beautiful.

So we can learn at least three lessons from thinking of the beautiful as the constitutive aim of aesthetic enjoyment: (i) the relation between the beautiful and aesthetic appreciation, call it $R$, implies that the beautiful is a proper object of aesthetic appreciation; (ii) that $R$ obtains implies that something’s being beautiful gives us good grounds for aesthetically admiring it; and (iii) that $R$ obtains imposes limits on what counts as aesthetic appreciation, at least such that if someone knows that some object (or feature of an object) is not beautiful, it is not possible for that person to aesthetically appreciate it.

I think (i)-(iii) state truths about the relation that obtains between aesthetic appreciation and genuine beauty, and that they are at least part of the story, not necessarily the whole of it, for explaining why aesthetic enjoyment of the beautiful forms an organic unity. If that is right, (i)-(iii) can be useful for ascertaining whether the relation between good motivation and objectively right action is of the sort that permits and conduces to the formation of organic wholes. I do not know whether satisfying (i)-(iii) is strictly necessary for the relation to be of the right sort, nor do I claim that it is always sufficient. I intend the more modest claim that, if the relation between good motivation and objectively right action satisfies (i)-(iii), that would give us significant

---

29 It is more complicated whether merely believing that some feature of an object is not beautiful is incompatible with aesthetically appreciating it. We can imagine a weak-willed aesthetic appreciator who believes that some object that is in fact beautiful is ugly, but cannot help admiring it in a way that might count as aesthetic enjoyment.
evidence for thinking that the relation between them is of the right sort to form an organic whole. *Does* the relation between good motivation and objective rightness, call it $R^*$, satisfy (i)-(iii)?

To say that $R^*$ satisfies (i) would be to say that the relation between good motivation and objective rightness is such that objective rightness *per se* is a proper object of good motivation: that doing the objectively right act is always an appropriate aim of the well-motivated deliberator. This is distinct from the idea that deliberation or practical reasoning *by its nature* aims at the right, or that we necessarily choose things *sub specie boni*. These latter claims are, of course, highly disputed. But even if they are false, it seems hard to deny that it is always appropriate to aim to do what’s right in light of the facts. *A fortiori*, an appropriate aim of the well-motivated person is to do what’s right in light of the facts. This is reflected in the fact that the well-motivated person does not just do what she takes to be best in light of her beliefs; she also hopes her beliefs are true, and takes reasonable steps to ensure that they are. While there might be cases where the well-motivated person takes something other than doing what’s objectively right as her end (while still counting as well-motivated), it seems undeniable that doing what’s objectively right is always at least consistent with being well-motivated. As I construe (i), that is enough to show that $R^*$ satisfies it.

To say that $R^*$ satisfies (ii) is to say that, if some act is objectively right, or has features that contribute to making it objectively right, this gives a well-motivated person reason to do it. This seems hard to deny, and indeed falls out of the fact that $R^*$ satisfies (i). Since doing the objectively right act is an appropriate aim of the well-motivated person, it is perfectly appropriate for well-motivated deliberators to take an act’s objective rightness or (objectively) right-making features as good reasons for action. Indeed some are likely to think that an act’s objective rightness or right-making properties give *anyone*, not just the well-motivated person, reason to do it. While I am sympathetic, this stronger claim is likely to embroil us in controversy: some deny that facts about rightness give everyone, rather than only those who have a standing concern with morality, reasons for action. Since I claim that it is only well-motivated objectively right acts that form organic wholes, I just need the weaker claim that facts about

---

30 For defense of the guise of the good hypothesis, Tenenbaum 2007. For a case against it, Stocker 1979. See also Tenenbaum 2010.

31 This is of course an enduring issue of Western moral philosophy. See e.g. Plato’s *Republic*; Prichard 1912; Foot 1972.
objective rightness give the well-motivated person reasons for action, in order for $R^*$ to satisfy (ii). Likewise, while it’s an interesting question whether objectively right-making features are the only good reasons for action, I do not need to settle it here. For $R^*$ to satisfy (ii) it is enough that objective rightness or right-making features are one appropriate source of reasons for action. This seems undeniable; $R^*$ satisfies (ii).

Finally, to say that $R^*$ satisfies (iii) is to say that facts about the objective rightness of acts impose limitations on what can count as instances of good motivation. I claimed above that, in the case of aesthetic enjoyment of beautiful, the beautiful imposes limitations on what can count as aesthetic enjoyment by implying that someone who knows that an object (or features of an object) are not beautiful cannot aesthetically appreciate that object (or those features). Similarly, knowledge of an act’s objective wrongness puts constraints on what counts as well-motivated deliberation. A well-motivated person could not pursue an act that she knows is objectively wrong, or take what she knows to be its wrong-making features as good reasons for action, while still qualifying as well-motivated. Indeed we can go further and say that, to be well-motivated, we must take our knowledge that acts are objectively wrong (or have objectively wrong-making features) as reasons against doing them. In this sense it seems clear that facts about objective rightness and wrongness put constraints on what can count as good motivation; so $R^*$ satisfies (iii).

What emerges from all this is a set of abstract structural similarities between the relation that characterizes aesthetic enjoyment and the genuinely beautiful ($R$), on the one hand, and the relation between good motivation and objectively right action ($R^*$), on the other. Insofar as enjoyment of the genuinely beautiful is a paradigmatic case of organic unity, the similarity between $R$ and $R^*$ provides evidence for thinking that $R^*$ (like $R$) is the sort of relation that permits and conduces to the formation of organic wholes. Like aesthetic enjoyment of the genuinely beautiful, the well-motivated performance of objectively right action is, at an abstract and general level, a good candidate for organic unity.

---

32 Some might deny that ‘because it is objectively right’ is something the well-motivated person appropriately takes as a reason for action, since this would entail that the motive of duty is always good, which it is not: sometimes it is fetishistic (Smith 1994). I address this in chapter five.
This said, there remain questions about $R^*$ that have the potential to put significant constraints on VA. In particular, it does not follow from the fact that the relation between good motivation and objectively right action is, at a general and abstract level, of the right sort to permit and conduce to organic wholes that there are not constraints on the relation that must obtain between motives and objectively right acts in particular instances. For all that’s been said so far, it could be that the relation that must obtain between $m$ and $s$ for them to comprise an organic whole puts limits on what counts as a good motive, in the sense relevant for VA.

We can put this as a worry about the scope of the organic unity approach. VA claims that any old admirable motive will do for virtuous action. And since VA states a necessary condition on virtuous action, the organic unity interpretation of VA implies that all virtuous acts are organic unities. It follows that any form of admirable motivation is capable of combining with permissible success to form organic wholes. But this might be too nonchalant. It could be that, for particular instances of admirable motives + permissible success to comprise organic unities, they must be connected in a highly particular sort of way—a way that secures a greater degree of unity than whatever unity is involved in satisfying (i)-(iii). If so, the form of unity required may put constraints on what motives count, on VA, as good.

In particular, it might be thought that particular instances of good motivation + permissible success count as unified in the right way only if the content of the motive from which the act is done, and the explanation for why the act is objectively right, are the same. Individuals must, on this view, perform objectively right acts because they are right, or for the reasons that make them right, to secure the relevant unity. If so, VA must either limit virtuous motivation to the desire to do acts for right-making reasons, or admit that not all virtuous acts are organic unities.

This is a challenging objection. I agree that right acts performed for right-making reasons are impressively unified. People motivated by the right-making properties of acts are attuned to what’s important about them, and do them for those reasons. Moreover, when people motivated}

---

33 I am grateful to Tom Hurka for this suggestion. Cf. Hurka 2001, chapter 1. For more general discussion of the relation between right-makers and good motivation see Arpaly 2003; Markovits 2010; Stratton-Lake 2000.

34 This would bring us closer to accepting the stronger constraint under (ii) above: that $R^*$ is of the right sort only if considerations having to do with an act’s objective rightness are the only good motives for doing the act. Again, see chapter five.
in this way succeed at doing the right thing, it is no mystery why. Desiring acts for the reasons that make them right is likely to be the most reliable ways actually to do right acts. Indeed it is difficult to see what other motive could produce a more reliable connection between good motives and permissible success. It thus seems true that objectively right acts performed for objective right-making features exhibit a high degree of unity, in a recognizable and intuitively attractive sense.

However, we should not just assume that entities must exhibit impressive forms of unity to form combinations with their own significance. This is a substantive claim in need of support, one that raises complicated issues. Perhaps we can get a handle on the issues by considering again Moore’s original paradigm of organic unity: the appreciation of true beauty. As I said earlier, appreciation is an intentional attitude directed at an object. Moreover, this directed-ness—the fact that you are appreciating this object—usually has an explanation: we typically appreciate things because they have, or we take them to have, certain features. The same is true of desiring or being motivated to act: we often, perhaps always, desire to do acts because they have, or we take them to have, certain features. Appreciating a beautiful object for the properties that make it beautiful is thus formally the same as desiring to do acts for the reasons that make them right. In both cases we have an intentional pro-attitude directed at an object for the features that make it a good or fitting object of the attitude in question.

The objection we are considering claims that well-motivated objectively right acts form organic wholes only when the motives involved are good in precisely this way: by having as their content whatever explains the objective rightness of the relevant acts. But now consider the parallel claim about appreciating beauty: that we must appreciate beautiful objects for their beauty-making features in order for there to be an organic unity. And think here of a beautiful painting, say, one of Van Gogh’s Sunflowers. What are the features that make it beautiful?

This is, for me (and many others I daresay), a daunting question. The beauty of Van Gogh’s Sunflowers seems to have something to do with the colors, the shapes, and the proportions. While perhaps true, this reply borders on vacuous and is of course extremely facile. Artists steeped in painting and art criticism will presumably have much more sophisticated and illuminating stories to tell about what explains the beauty of Van Gogh. But then what level of abstraction matters for whether appreciation of the Van Gogh is an organic unity? Must we
appreciate it for the naïve and arguably vacuous explanation (because of its colors, shapes, and proportions), or must we appreciate it for more sophisticated reasons (because the colors, shapes, etc. are combined thus-and-so)? It seems we should not require appreciation to track the more precise beauty-making story, since this would disqualify most cases of appreciating beauty as organic unities. Presumably for most people the appreciation of beauty is not based on a sophisticated grasp of aesthetics.

Just as there are more and less determinate ways of specifying beauty-making properties, there are more and less determinate ways of specifying right-making properties. Suppose that your friend, Steve, is sad. Saddened by Steve’s sadness, and wanting to cheer him up, you give him a puppy. Alas, Steve does not particularly like dogs; indeed he mildly dislikes them. His daughter, however, is thrilled by the puppy. In fact Steve’s daughter is so happy that her happiness makes giving Steve the puppy the objectively right thing to do. Since it is also true that you gave Steve the puppy from admirable motives (to cheer up Steve), we have here a case of well-motivated objectively right action.35 But since what moves you in this case is a desire to make Steve happy, and since it is Steve’s daughter’s happiness (not Steve’s) that makes the act right, it might seem that you do not give Steve the dog for right-making reasons—in which case, according to the objection, your act does not form an organic whole.

However, whether you give Steve the puppy for right-making reasons depends on further facts about your motives. While giving Steve the dog is not right because it makes Steve happy, it is right because it makes someone happy. And since this is true, it is also true that giving Steve the dog is right because it accomplishes something good. Moreover, if we take ‘desiring acts for right-making features’ to include the motive of duty (as we should), desiring acts because they are right will also count as desiring them in the right way.36 Whether you give Steve the dog for

35 If you are inclined to doubt that your motives are really good in this case, on the grounds that someone well-motivated would make sure that Steve likes puppies before giving him one, we can fill in the case so that by all appearances Steve likes puppies, in such a way that it is reasonable for you to conclude as much, even if in reality he does not.

36 Some deny that an act’s rightness can be one of its right-making features. See Dancy 1993; Stratton-Lake 2000. Even if this is true, desiring right acts because they are right seems to involve a high degree of unity between motive and rightness, which is what the objection insists upon. The view of the authors above might make us restate the unity requirement in terms other than ‘right-making reasons,’ but should not keep us from affirming that the motive of duty meets any plausible unity requirement.
right-making reasons depends on whether you desired to give Steve the dog under any of these less determinate descriptions. To insist that you must give Steve the dog to make Steve happy—rather than because it makes someone happy, or does some good, or is the thing to do—seems analogous to requiring the appreciation of beautiful objects to be based on highly determinate descriptions of their beauty-makers. In both cases we risk ruling out plausible cases of the phenomenon we wish to capture.

If desiring right acts under these less determinate descriptions counts as desiring them for right-making reasons, it’s no longer clear how many cases the objection rules out—even if we agree that acts must be desired for right-making reasons—since often acts will be desired under one or another of these descriptions. While it is true that you desire to give Steve the dog to make Steve happy, it may also be true that you desire to give Steve the dog because it will make someone happy, and/or do some good in the world, and/or is right. This is particularly likely if we think that our dispositional desires or reasons for action, indicated in post-hoc answers to ‘Why?’ questions, are relevant for determining the content of motives. Suppose you ask me why I gave Steve the dog. I might first say: ‘To make Steve happy.’ You then reply: ‘But why did you want to make Steve happy?’ While it doesn’t follow from the fact that I gave Steve the dog to make Steve happy that I gave Steve the dog to make someone happy, or do good in the world, or do the right thing (since the context is opaque), it would not be at all surprising for me to reply: ‘Well, just to make someone happy; to do some good in the world; because it was right.’ It is, I think, typically true that someone who desires acts under relatively concrete descriptions often desires them (perhaps dispositionally) under these more generic descriptions as well.

But the main point is that, just as it is implausible to insist that beauty must be appreciated under highly determinate descriptions to count as an organic unity, the same holds for the combination of good motives + permissible success. This suggests that the degree of unity required may be weaker than the objector is inclined to think.

To further press this point, consider the suffering of the vicious. Many have held that it is a good thing that the wicked suffer.\textsuperscript{37} This is difficult to explain without the concept of organic unity,

\textsuperscript{37} For the record, I myself do not have a strong sense that the suffering of the vicious is good. But many impressive philosophical lights do seem to accept this, among them Kant, Sidgwick, Moore, and Ross.
since considered on their own both the viciousness that makes people deserving of suffering, and the suffering itself, seem intrinsically bad. If the combination of the two is per se good, this must be because they form an organic whole. What sort of unity must viciousness and suffering exhibit for their combination to be good?

The unity between viciousness and suffering seems to be secured by the fact that the vicious deserve to suffer. But now consider different ways in which the vicious might suffer: different sources of their suffering.

Imagine that Stan is vicious toward his family and colleagues. He is always abusing them verbally, completely ignoring the effects of his actions, and generally being an insufferable jerk. Suppose now that Stan begins to suffer in ways proportional to his vice, but that the sources of Stan’s suffering are in no way connected to his viciousness. Stan just begins to have really bad luck: through no fault of his own he contracts various ailments, loses large amounts of money on the stock market, and so on. While Sam deserves to suffer, it’s not clear that he deserves to suffer in this way. Sam deserves to suffer because he is vicious. This ‘because’ could denote either a causal relation, such as where some wicked choice of Sam’s sets in motion a chain of events that results in his suffering; or (if they differ) an intentional relation, such as when someone who imposes suffering on Stan is motivated by the thought of his viciousness. Stan’s suffering is brought about in neither of these ways. Nonetheless, we might still think that Stan’s suffering is good. We are commonly pleased when blind fate ruins villains in movies and novels. While it may be more fitting for Stan to be undone by the consequences of his own deeds, or for those who punish the vicious to be motivated by their viciousness, this is not necessary for Stan’s suffering to strike us as good. Insisting on a high degree of unity here would preclude what many will see as a genuine organic whole.

Or consider again cases of non-traditional organic wholes: combinations of moral factors whose significance is not, or need not be, explained in terms of intrinsic value. If, as Kagan suggests, doing harm is generally less bad when defending one’s self than in other contexts, must there be some sort of unity between doing harm and self-defense in order to explain this? It seems not. What would such a unity look like or be?
Likewise, consider Kamm’s suggestion that the badness of letting someone die depends on how much personal sacrifice would be required to save her. Is there some unity needed between great personal sacrifice and letting die for letting die to be less bad? Again, it is hard to see what such a unity would even come to, never mind that it is required.

Perhaps the most that someone who insists on a high degree of unity can say about these cases is that there is something intuitively fitting about the idea that doing harm is less bad when one is defending one’s self; and something intuitively fitting about letting die being less bad when great sacrifice is required. But the same can be said of the claim that well-motivated objectively right acts are per se significant. It is intuitively fitting to think that motives matter to the value of permissible success, and vice-versa. Indeed in light of the fact that $R^*$ satisfies (i)-(iii), the degree of unity that exists, at a general level, between good motivation and objective rightness is likely to be significantly greater than the degree of unity that exists between these non-traditional cases of organic wholes.

Finally, our intuitions about cases that isolate the relevant difference do not clearly support the conclusion that objectively right acts must be done for the reasons that make them right in order to comprise organic wholes. It also seems to me that these intuitions are not idiosyncratic or baseless, but can be given a satisfactory grounding.

Consider again your attempt to raise Steve’s spirits. The objection I’ve been considering insists that this is not an organic whole, since there is not enough unity between your motives and the act’s rightness: your reasons for giving Steve the puppy (to make Steve happy) are not what actually make the act right (Steve’s daughter’s happiness). However, it’s far from obvious that making Steve’s daughter happy is irrelevant to the significance of your act. It certainly seems better that Steve’s daughter likes the puppy (thereby making your act right) than that she does not (thereby making it wrong), even if this is due to sheer luck.

Moreover, however lucky you may be, it is nonetheless true that by giving Steve the dog you make Steve’s daughter happy, and that you are well-motivated in doing so. The causal role you play here seems evaluatively significant. Imagine that you fail to give Steve the dog (it breaks its leash and escapes as you are walking up to Steve’s house), but that Steve’s daughter finds a runaway puppy that same day, which she insists on adopting, and who makes her just as happy.
as she was in the original case. I think many of us would prefer the original case, in which we
make Steve’s daughter happy, to this one.\(^{38}\) We would rather succeed at doing good, even when
we are lucky, than to try but fail to do good, even if no good is lost.

This intuition can, moreover, be given a plausible grounding. Being the cause of good things,
even when this happens in ways we do not fully predict or intend, can be seen as having
significant positive import. In general, the Western philosophical tradition has been hostile to
the thought that being a mere cause of goodness—rather than being an agent who \textit{intends}
goodness, or intends it for the right reasons—is valuable. Apart from some (but not all)
consequentialists, it is difficult to find defenders of the idea that merely causing good is
evaluatively significant.\(^{39}\)

Interestingly, the same is not true of the idea that merely causing \textit{bad} is evaluatively significant.
In their famous exchange on moral luck, for example, Bernard Williams and Thomas Nagel both
agree that being a (mere) cause of evil is bad.\(^{40}\) The Williams-Nagel exchange (and much of the
debate it spawned) focused on whether being the cause of evil is bad in some distinctly moral
sense. But whether or not morally bad, it seems very plausible that being the cause of bad makes
\textit{one’s life} worse. Consider Williams’ famous example of the lorry driver who, non-intentionally
and through no fault of his own, runs over a child.\(^{41}\) This constitutes a splotch on the driver’s
life. Williams captures this thought by noting that a particular kind of regret, ‘agent-regret,’ is
appropriate. While merely causing bad may not detract from the goodness of lives as much as
intentionally causing bad, or be as bad for us as other ways in which our lives are made worse
(such as physical or emotional suffering, or having our deeply held desires frustrated), we prefer

\(^{38}\) We may also regret that Steve doesn’t like the puppy. But this is compatible with preferring,
all-things-considered, to have caused Steve’s daughter’s happiness.

\(^{39}\) Objective consequentialists who make rightness (or virtue) a function of goodness actually
produced recognize the importance of merely causing good; it is less clear that the same can be
said for subjective consequentialists who make rightness (or virtue) a function of expected value.
To my knowledge, the only defense of this claim outside of an explicitly consequentialist
framework is Wendler 2010, ch. 7. I am indebted to that discussion in what follows.

\(^{40}\) Williams 1976 and Nagel 1976.

\(^{41}\) Williams 1976: 124.
not to be the cause of evil. The best way to explain this is that being the cause of bad makes our lives worse, to at least some degree.  

If we accept that merely causing bad makes our lives worse, it is natural to think that merely causing good makes them better. This does not follow logically, since there may be an interesting asymmetry between the significance of causing bad and the significance of causing good. But the idea that causing bad makes our lives worse ought at least to create a presumption that causing good makes them better.

It seems to me that this presumption holds up. Consider, for example, what denying it implies about causal episodes involving children too young to understand the significance of their acts. It implies that we have no grounds to prefer having been a child who consistently caused others happiness and joy to being a child who (through no fault of his own) caused pain and suffering, say, by being naturally ill-tempered. This seems to me dubious. If given a choice, many of us would, I wager, prefer to live a life where our childhood was spent causing joy to a life where our childhood was spent causing suffering, even if in neither case we are responsible for it.

The view that causing good makes our lives better also helps to explain why it can be appropriate for parents to place children in projects that contribute to good aims, even if the children do not understand their role in them. Consider young children who participate in clinical research aimed at alleviating debilitating childhood conditions. It seems plausible that the contributions made by these children make their lives better, in some important sense. Someone who as a child contributes to pediatric research might look back on this with a sense of pride; they might value the fact that they made a causal contribution to a worthwhile project, even if they did not at the time understand what they were doing. This would, I think, be entirely appropriate. Similarly, we might be appropriately quite pleased that giving Steve the puppy made his daughter happy, and so was objectively right, even if this was not how we understood the rightness of the act when we did it.

---

42 It is not necessary to hold that merely causing bad detract from our welfare, since it is possible to hold that welfare is only one scale among others on which the goodness of lives can be weighed. Causing bad may make lives worse on one of these other scales. Mutatis mutandis for causing good.

43 Cf. Raz 2004: 74, who asserts without argument that there is such an asymmetry.
While the claim that being a mere cause of goodness improves our lives needs more defense, the preceding considerations suggest that our intuitive preference for causing objectively good ends can be rationally grounded. If so, it can be better to be well-motivated and inadvertently realize permissible ends than to be well-motivated and fail to realize them, even if the ends are brought about in some other way. This shows that the value of well-motivated attempts to realize good ends and the value of realizing good ends cannot be understood in isolation, even when the realization of the end is lucky. This in turn rebuts the objection: well-motivated objectively right acts need not be done for the reasons that make them right for them to constitute organic wholes.

1.4.5. *Is permissible success necessary?*

Yet another objection to the account so far focuses on the objective rightness component of VA, questioning whether it is really necessary for virtuous action. In one of its guises the objection advances cases of seemingly virtuous action that are not objectively right. I will postpone this form of the objection until next chapter. For now I want to look at what I think is a more interesting challenge, one focused more directly on the claim that good motives + permissible success forms an organic unity.

The objection I have in mind concedes that well-motivated objectively right acts constitute organic wholes, but denies that it is facts about objective rightness that explain this. Instead it is some other fact that explains it. If realizing an end that isn’t objectively right (from good motives) can form the same sort of organic whole as realizing an end that is objectively right (from good motives), the organic unity interpretation of VA fails to say what’s distinctive about virtuous acts.

One sort of example the objection could use looks like this:

*Helping Bill*

Imagine that you are helping your next-door neighbor, Bill, move some furniture, because you want to do something nice for him. At the same time your other neighbor, Fred, is choking in his home, which is right next to Bill’s. Being an expert in the Heimlich maneuver, you could easily save Fred, if you knew he was choking. But you do not, and Fred dies.
This works as a counterexample by assuming that helping Bill is, under the circumstances, objectively wrong, since Fred is choking and you could easily save him. If helping Bill is objectively wrong, VA entails that helping Bill from admirable motives does not form an organic whole, since on VA only virtuous acts form wholes in this way; and virtuous acts require objective rightness.

Intuitively, however, no less here than in the cases used above (Mandela, etc.), the value of your motive matters for the value of successfully helping Bill, and/or vice-versa. Your attempt to help Bill seems more valuable when it succeeds, and/or the success seems more valuable when it is well-motivated, even if helping Bill is objectively wrong. In cases like this we have a non-virtuous act combining good motives with objectively impermissible success, apparently to form the sort of organic whole that virtuous acts alone are supposed to form.

While this worry for VA is real, I don’t think cases like Helping Bill are the best ones to use in the objection. For it’s not clear to me that helping Bill really is objectively wrong, in any sense that proponents of objective rightness and wrongness should care about. The conclusion that helping Bill is objectively impermissible depends on the assumption that helping Fred is objectively required. But there is what seems to me a plausible argument against this thought. The argument is simple, though it points to deeper issues.

P1. Act a is objectively required of you, in the sense we care most about, only if it is reasonable to expect you to consider a-ing.

P2. It is not reasonable to expect you to consider helping Fred.

C. Helping Fred is not objectively required of you.

In a nutshell, I think we should accept P1 because it does better justice to the functional role of ‘objective rightness’ in deliberation and action-guidance; better, that is, than views on which ‘the objectively right act’ is specified from all of the myriad and often overwhelming number of acts we merely could or can do. And I think we should accept P2 because it is not reasonable to expect people to consider doing acts for which they have no evidence that they should do (and, perhaps, provided they are not negligent in their evidence-gathering). But properly defending both of these ideas would require a lot of argument. Since there are other counterexamples that
do not assume that we can be objectively required to do things that it is not reasonable to expect us to consider doing, I’ll consider one of those instead.\(^{44}\)

*Saving One*

You are on the lake in your canoe, when you hear shouts from both your left and your right. To your far left, on the edge of the horizon, five people are flailing about in the water, calling for help (you know there are five because you have binoculars). To your right, closer to the boat but still a fair bit off, one person is in the water calling for help. You judge, correctly, that you do not have time to save all six. By the time you save the five the one will drown, and vice-versa. Because the five are so much further away, you are uncertain whether you will be able to make it to them before they drown; whereas you are confident that you can make it to the one in time. You paddle over to the one and save him; the five drown. However, as a matter of fact, you really did have time to save the five. Had you paddled over to them, you would have saved them.

Suppose now that, given this description of the case, saving the five is objectively required of you, and saving the one is objectively wrong.\(^{45}\) Suppose also that you are well-motivated in saving the one. This would require you to do so out of concern for the one. It would also require your deliberative error—your judgment that you do not have time to save the five—to be compatible with all-things-considered good motivation. But both of these seem entirely possible; suppose they obtain.

Given this description of the case, the organic unity approach predicts that saving the one does not form an organic whole, since it is objectively impermissible. Intuitively, however, saving the one from good motives might seem a good candidate for an organic whole. Your attempt to save the one may seem more valuable when it succeeds, and/or actually saving the one may seem better because it is well-motivated, even if saving the one is objectively wrong. If so, *Saving One* will be a counterexample to the organic unity interpretation of VA.

\(^{44}\) I raise the *Helping Bill*-type case because both Tom Hurka and Sergio Tenenbaum proposed counterexamples very close to it.

\(^{45}\) Not everyone, of course, will accept this; some will insist that what is objectively right depends on the outcome of your flipping a (perhaps weighted) coin. These people will have to find a different example.
The strength of this challenge depends on the intuition that saving the one rather than saving the five forms an organic whole. Since the objection will be much stronger if it can plausibly ground the intuition to which it appeals, we should, I think, ask why saving the one is an organic whole. If the objection struggles to answer this question, there will be room for skepticism about it.

What could explain why saving the one is an organic whole? It would be strange if it were impermissibility itself that explained the boost in value. Instead I take it that Saving One gets whatever force as a counterexample it has from the fact that saving the one is a good outcome: in acting as you did you have preserved a thing of great value—the one’s life—even if doing so was objectively wrong. It is, on this suggestion, the goodness of the outcome or state of affairs realized—the preserving of the one’s life—that combines with your good motives to form an organic whole.

One reason for denying that bringing about a good outcome per se is the relevant fact is that the relation between good motivation and good outcomes is dis-analogous, in at least two important ways, from the relation that characterizes other organic wholes involving intentional attitudes and intentional objects. In particular, the relation between good motivation and good outcomes, call it $R^{**}$, does not resemble the relation between aesthetic appreciation and the genuinely beautiful ($R$) to the same extent that the relation between good motivation and objective rightness ($R^*$) does. Earlier I argued that $R^*$ resembles $R$ by virtue of satisfying (i)-(iii), and that this gives us evidence for thinking that $R^*$ is of the right sort to characterize organic wholes. $R^{**}$, by contrast, is likely to fail on both (i) and (iii).

For $R^{**}$ to satisfy (i) it would have to be true that bringing about good outcomes per se is always an appropriate aim of the well-motivated person. While it may be true that the well-motivated person often aims to act in ways that have good consequences, this is because, and to the extent that, doing what’s right has good consequences. But it is not always good to aim at good consequences. When bringing about good consequences is wrong, and we know this is so, aiming to do so can be inconsistent with good motivation. Aiming to do an objectively permissible act, by contrast, seems always consistent with good motivation.
For \( R^{**} \) to satisfy (iii) it would have to be true that facts about good consequences constrain what count as instances of good motivation, such that someone who knows that an act has bad consequences cannot pursue this act while continuing to count as well-motivated. This again seems false, for reasons similar to above. Sometimes doing what’s right might involve bringing about bad consequences.\(^4\) In these cases the well-motivated person can pursue what is right, even if it involves bad consequences, consistent with her being well-motivated.

Given that \( R^{**} \) fails to satisfy (i) and (iii), we can, I think, at least say that \( R^{**} \) is not as strong a candidate for the right sort of relation, the relation characteristic of organic wholes, as is \( R^{*} \), the relation between good motives and objective rightness. On a purely abstract level, \( R^{**} \) is outperformed by \( R^{*} \).

Apart from these more abstract considerations, the proposed analysis of *Saving One* is subject to difficulties. The proposal is that saving the one forms an organic whole because, by saving her, you realize a good state of affairs. Being well-motivated and bringing about a good outcome suffice, on this view, for an organic whole. As a general principle, however, this seems false.

Suppose, for example, that preventing evil requires an egregious rights-violation. You must choose to push the fat man off the trolley bridge, or let the trolley hit and kill two bystanders. Suppose we fill in the case so that you are well-motivated in pushing the fat man. You are deeply concerned for the two on the tracks; having been raised in a staunchly utilitarian home, you think that killing the man is permissible and indeed required; and so on. In that case killing the one to save the two should form an organic whole. Your attempt to kill the one should be better because it succeeds, and/or the fact that you are well-motivated should make killing the man, considered on its own, better. But this is, I think, likely to strike many of us as false. I doubt we want to regard our attempt to kill the fat man more favorably because it succeeds, and the fact that your motives are good does not seem to make the killing *per se* any less bad.

More cautiously, whether we find these ideas plausible may well depend on whether we think killing the one is the objectively right thing to do. If we think pushing the fat man is permissible, we will, I suspect, be much more inclined to think that there is an organic unity present. Dyed-

---

\(^4\) This is likely to be true even for consequentialists, for example, in cases where preventing great evil requires producing lesser evil. There are questions here about what should count as ‘good consequences’; see below.
in-the-wool consequentialists who think that pushing the fat man is right are likely to focus on the rightness of saving two lives, in a way that makes it plausible to accept that there is an organic whole. By contrast, deontologists who think that killing the one is wrong are likely to focus on the nature of the rights-violation, in a way that vitiates the idea that killing the one interacts valuably with the content of the motive (even if it’s ATC good). Arguably, then, this case can be seen as supporting the idea that objective rightness, over-and-above bringing about good outcomes, is needed for well-motivated acts to form organic wholes.

If, as deontologists are likely to conclude, bringing about good outcomes from good motives in this case does not suffice for an organic whole, we should be skeptical that good outcomes explain the existence of the organic whole in Saving One. On the other hand, as we have just seen, cases like this will not be useful for dislodging consequentialists of the idea that good outcomes per se are what enters into the relevant wholes, since for consequentialists the case does not isolate the difference between good outcomes and objective rightness. But there are other cases that do isolate this difference that seem to cut against the idea that good motives + good outcomes suffice.

Suppose again that a runaway trolley is headed toward two bystanders, and that you are standing on the footbridge next to the fat man. Imagine that there is also a large barrel on the footbridge next to you. The barrel is big, though not as big as the fat man. You must, you think to yourself, either toss the barrel on to the track, and hope that it is big enough to stop the trolley; or push the fat man on to the track, whom you are certain is big enough to stop it. Because you are sure the fat man will stop the trolley, and unsure whether the barrel will, you push the fat man. Suppose further that your motives in doing so are ATC good: you push the fat man out of concern for the two; you have been raised in a staunchly utilitarian home; and so on. Unfortunately, you make the wrong choice: the barrel would have sufficed to stop the train.

In this case even utilitarians can agree that killing the fat man is objectively wrong. At the same time, killing the fat man has good consequences, since it prevents the death of the two.

47 Of course, killing the fat man does not have good consequences when compared with throwing the barrel. But by the same token saving the one in the original counterexample does not have good consequences when measured against saving the five. I see no way to deny that killing the fat man has good consequences without also denying that saving the one has good consequences.
bringing about good consequences from good motives suffices for an organic whole, killing the fat man from good motives should form an organic whole. This again is likely to strike us as false.

It seems to me that these cases work as counterexamples to the idea that good motives + good outcome suffices for an organic whole. It’s worth noting that they also cut against a second possible explanation for what could explain an organic unity in Saving One, namely, the idea that well-motivated (merely) subjectively right—not necessarily objectively right—action forms an organic whole. Given your beliefs in Saving One (that you do not have time to save the five, and so on), saving the one is likely to be subjectively right. Perhaps, then, your good motives interact with the subjective rightness of saving the one in a way that can explain the existence of an organic whole.

But if well-motivated (merely) subjectively right action suffices for an organic whole, pushing the fat man in both cases—when the barrel is present and when it is not—should be an organic whole. In both cases you are, after all, well-motivated, in a way that makes it easy to stipulate that pushing the man is right from the perspective of your beliefs. You believe that pushing the man is the only way to stop the trolley; that it is permissible to push him; and so on. No less than when good consequences (in conjunction with good motives) is the proposed explanans, these acts should, if subjective rightness is the explanans, form organic wholes. But since they are the very same cases, any intuition we had above that pushing the fat man does not form an organic whole applies here as well. So it seems to me that these cases also undermine the idea that good motives + subjective rightness suffices for an organic unity.48

We thus have two failed attempts to ground the intuition that saving one in Saving One is an organic whole. Moreover, these two strategies seem by far the most promising. If so, their failure should significantly erode our confidence in the intuition on which the objection is based. There is, in addition, what seems to me a plausible debunking explanation for any sense we might have that saving the one in Saving One forms an organic whole. This intuition can be explained simply by our tendency to associate saving a life with doing something that is in fact

48 I would also argue that the relation between good motivation and subjective rightness is dis-analogous to the relation between \( R \) and \( R^* \), by virtue of failing to satisfy at least (i) and (iii) above. But this would take a lot of argument.
permissible. In ordinary contexts it is rarely ever wrong to save a life. Thus, it will rarely ever fail to be the case that saving a life from good motives forms an organic unity. In this light our initial gut reaction to Saving One is natural and perfectly understandable, even if, as I have argued, it is very difficult to ground.

Where does this leave us? Even if VA must debunk intuitions about cases like Saving One, the preceding analysis shows how other possible explanations for the intuition on which the objection depends—that saving the one forms an organic whole—will have to do at least as much explaining away, and likely more. It seems to me that this neutralizes the objection.

1.5. VA motivated

Having now developed VA in what I take to be a theoretically and intuitively attractive way, we can return to the charge that VA is ill-motivated. The sophisticated P-R objector claimed that VA is ill-motivated, insofar as it takes two evaluative factors that everyone already agrees are operative in act evaluation and applies the term ‘virtue’ to their conjunction. We can now see that this objection is misplaced. While VA applies ‘virtue’ to a conjunction, the combination it picks out has its own evaluative significance, in addition to the significance of each of its conjuncts. To put it somewhat metaphorically, VA recognizes the special value that exists in well-motivated objectively right action, considered as a unified whole, and uses virtue-terms to capture this value. Given that well-motivated objectively right action constitutes its own evaluatively significant whole, the objection that VA is a ‘merely conjunctive’ theory of virtuous action has no force.

This said, it’s also not obvious that the organic unity interpretation gives us reason to prefer VA to P-R. For it remains true that P-R, like VA, acknowledges the existence of well-motivated objectively right action, even if P-R places it beneath a different label, ‘overall superlativeness.’ It might seem that proponents of P-R can adopt the organic unity analysis of well-motivated objectively right action while continuing to claim that well-motivated objectively right acts belong under the label of ‘overall superlative,’ rather than ‘virtuous,’ action.

There is, however, a serious difficulty with attempting to assimilate VA’s approach in this way. Suppose that defenders of P-R, while denying that objective rightness is needed for virtuous action, accept that well-motivated objectively right acts form organic wholes. Now take some
act that satisfies this description, say, the act of meeting Kim at the station, which constitutes the keeping of a promise, from (say) the motive of duty. P-R says that this act is virtuous, since (ex hypothesi) it is done from a good motive. Moreover, P-R claims that the virtue of this act is a function solely of the motive from which it is done. But now P-R faces a dilemma, one that arises from accepting that well-motivated objectively right acts form organic wholes. Either the objective rightness of meeting Kim at the station bumps up the value of your motives, as the conditional approach to organic unities says; or the extra value resides in a whole constituted by admirable motives + permissible success, as the Moorean interpretation claims.

Suppose first that the conditional approach is correct. It follows that the value of your motives depends on it being objectively right to meet Kim at the station. But if the value of your motives depends on facts about objective rightness, any view that makes the virtue of acts depend on motive implies that their virtue depends on objective rightness as well. If the virtue of acts depends on the value of motives, and if (as the conditional approach implies) the value of motives depends on facts about objective rightness, then insofar as ‘depends on’ is transitive, the virtue of acts depends partly on objective rightness. Since P-R denies that the virtue of acts depends on objective rightness, this is a problem for defenders of that view.

Suppose on the other hand that the Moorean approach to organic wholes is the right way to think about well-motivated permissible success. Then the value of motives and the value of permissible success remain constant, and the extra value is located in a whole constituted by the two. P-R can now say that the virtue of well-motivated permissible success depends just on motive, since the value of motives no longer depends on facts about objective rightness.

At the same time, P-R must on this approach admit that the significance of motives outstrips the role they play in making acts virtuous. While this approach can recognize that the virtue of acts depends just on the value of motives, it cannot affirm that the value of motives is fully reflected in the virtue of acts. This is because, on the Moorean approach, part of the significance of motives is the role they play in co-constituting and making valuable the relevant wholes: those

---

49 It might be thought that all that follows is that the degree of goodness of your motives depends on facts about objective rightness. But if the conditional approach to organic wholes is the right way to conceptualize VA, it is possible in principle for motives that would have otherwise been neutral (or perhaps even slightly bad) to be made good by facts about objective rightness.
consisting in admirable motives + permissible success. It is clear that proponents of P-R should resist making the value of these wholes, considered as wholes, relevant for the virtue of acts. If it were relevant, P-R would again need to admit that facts about objective rightness are relevant for the virtue of acts, since the value of the whole depends partly on facts about objective rightness.

If on the other hand the value of these wholes per se is not relevant to the virtue of acts, then the importance of motives is not fully reflected in an act’s virtue, since on the Moorean approach part of the significance of motives is the role they play in making the wholes valuable. While this horn of the dilemma seems less bad for P-R, it is still undesirable. It forces defenders of P-R to say that motives contribute to the evaluative significance of acts along two distinct axes: motives determine an act’s virtue, and also partly determine an act’s overall superlativeness (or whatever we want to call the value of the whole). This is messy and so far forth unattractive. It also looks unprincipled. Why should the significance of motives be partly reflected in an act’s virtue, and partly reflected in this other, completely unrelated way? It is much more theoretically appealing to hold that the significance of motives is fully reflected in the virtue of acts.

VA avoids the dilemma just sketched. By explicitly making the virtue of acts depend on motive and objective rightness, it has no trouble with the implication that the goodness of motives, and so the virtue of acts, depends on facts about objective rightness. VA can also recognize (should the Moorean approach be true) that the virtue of acts depends on the value of the whole consisting in admirable motives + permissible success. These considerations suggest that, contrary to first appearances, P-R will have difficulty capturing the thought that admirable motives + objective rightness form organic wholes. Insofar as that claim is attractive, and accepting it commits VA to unattractive views about the evaluative significance of motives, we have reason to prefer VA.

It seems to me that the organic unity approach to virtuous action gives VA another advantage over P-R. Once we see that good motives + objective rightness is an organic whole, we can see that VA fits better with a plausible view about the metaethics of virtue, namely, that virtue is a value that makes something an exemplary instance of its kind. If that is right, virtue terms function to pick out such excellences. Given that a well-motivated objectively right act forms a whole with its own distinctive value, and that the value of this whole surpasses the value
involved in merely well-motivated acts, the acts VA picks out will be better candidates for virtuous acts than the acts P-R picks out. An act that is not objectively right will, after all, lack whatever independent value is involved in permissible success, as well as the value that accrues from combining good motives + permissible success. By making the virtue-status of acts depend just on motive, defenders of P-R seem to be missing the evaluative significance of permissible success. Whereas the acts deemed virtuous by P-R excel in one area of excellence—along the motivational axis—the acts VA deems virtuous excel in two—along both the motivational and the rightness axis. Other things equal, acts that excel along both dimensions seem more exemplary than acts that excel along one. On the assumption that virtuous acts are particularly exemplary instances of their kind, we have reason to prefer VA to P-R.

Finally, the organic unity interpretation of VA yields an independently plausible relation between admirable motives and rightness. Some extreme views claim that rightness depends on being well-motivated, so that acts are right just in case and because they are well-motivated. As I argue in chapter four, these views return unacceptable conclusions about the deontic status of particular acts, and should be rejected. On the other end of the spectrum we find views, like P-R, that completely sever the significance of motive and rightness. P-R claims that the rightness of acts never depends on motive, and likewise that the value of motives never depends on their issuing in right acts.

As I have argued, however, this latter claim is implausible. While VA like P-R denies that rightness depends on motive (see chapter four), it shows how the significance of good motives cannot be fully understood apart from facts about rightness, and likewise gives us a way to capture the thought that motives, while not deontically relevant, matter for the overall evaluative significance of right acts. VA captures the intuitive thought behind motive-based theories of rightness—the common sense intuition that there is some evaluatively significant connection between motives and rightness—without going to the extreme of making rightness depend on motive. In doing so it carves out an attractive middle ground between motive-based theories of rightness and P-R.

I’ve been motivating VA by developing the idea that well-motivated objectively right acts have their own integrity and value, by virtue of forming organic wholes. While these claims about organic unity will continue to be relevant, I want now to move away from defending the thesis
that VA picks out organic wholes and begin to focus on other implications of the idea that objective rightness is necessary for virtuous act.
Chapter two
Further support

2.1. Virtue and subjective rightness

Recall that my primary aim in the first two chapters is to trace the implications of VA in a way that allows us to compare it with other views about virtuous action, and P-R in particular. One way VA differs from other views is in the conclusions it returns about the virtue-status of particular acts. A second way is in the relations VA supports between virtue and other moral concepts: especially the relation between virtuous action and subjective rightness, and the relation between virtuous action and the good of individuals and communities. In both cases what explains the differences is that VA does, and other views do not, make permissible success necessary for virtuous action. I will start with the relation between virtuous action and subjective rightness. After that I’ll consider VA’s implications for the relation between virtuous action and benefit, and conclude the chapter with an assessment of particular cases.

With respect to the relation between virtuous action and subjective rightness, VA implies two things. First, it implies that subjective rightness is not necessary for virtuous action. This follows from VA’s claim that good motivation + objective rightness is sufficient for virtuous action, in conjunction with two plausible assumptions: that objective rightness does not entail subjective rightness, and (less obviously but still I think plausible) that being well-motivated does not entail doing what’s subjectively right (see below.) Second, VA implies that good motives + subjective rightness is not sufficient for virtuous action. This follows from VA’s claim that both good motives and objective rightness are necessary for virtuous action. Well-motivated subjectively right acts must also be objectively right, on VA, if they are to be virtuous. Are these implications about the relation between virtuous action and subjective rightness plausible?

At a general level, assessing the relation between virtuous action and subjective rightness is complicated by the fact that ‘subjective rightness’ is a theoretical construct open to a variety of specifications. The basic idea is that subjectively right acts are right from the perspective of the agent, rather than the perspective of the facts. But there are many ways to understand ‘the perspective of the agent.’ We might identify ‘the perspective of the agent’ with all of her beliefs;
or all of her empirical (non-moral) beliefs; or all or some of her credences; or her evidence; or all or some of her justified or rational beliefs; and so on. Each of these gives rise to a different sense of subjective rightness and returns (or may well return) different conclusions about which acts are subjectively right.¹

It would obviously be a mammoth undertaking to argue that no possible sense of subjective rightness identifies a necessary condition on virtuous action. In what follows I limit my sights to two currently prominent notions of subjective rightness. The first takes subjective rightness to be a function of an agent’s factual and moral beliefs: the subjectively right act is the act best supported by all of the agent’s beliefs.² Since on this view the moral beliefs that partly determine the subjectively right act may be false, we can call this non-moralized subjective rightness. The second view takes subjective rightness to be a function of an agent’s factual beliefs in conjunction with the correct moral principles: the subjectively right act is the act the agent’s factual beliefs, taken together with the correct moral code, instruct him to perform.³ I will call this moralized subjective rightness.

So we have two questions to pursue. First, does either moralized or non-moralized subjective rightness supply a necessary condition on virtuous action? Second, does doing either a moralized or non-moralized subjectively right act from ATC good motivation suffice for virtuous action?

2.1.1. Sufficiency and subjective rightness

Consider the second question first. The same cases used above to undermine the idea that good motives + good outcomes suffice for an organic whole can be used to show that good motives + non-moralized subjective rightness do not suffice for an organic whole. Imagine again that, due to empirical uncertainty, you throw the fat man instead of the barrel to stop the trolley, when in fact the barrel would have sufficed. Above we filled in this case so that your throwing the fat man is both ATC well-motivated and supported by your moral and non-moral beliefs; in the terminology just introduced, throwing the fat man is non-moralized subjectively right. If good

¹ See Parfit 2011; Sepielli 2012.
³ F. Jackson 1991. Smith 2010 focuses her illuminating discussion on this concept of subjective rightness.
motivation + (non-moralized) subjective rightness sufficed for virtuous action, throwing the fat man instead of the barrel would be virtuous.

But just as killing the fat man instead of throwing the barrel from good motives does not appear to form an organic whole, it also does not appear to be virtuous. Or, more carefully, if it does strike some of us as virtuous, that is likely due to the fact that you act from good motives, rather than because your act is (non-moralized) subjectively right. VA’s response to cases like this is to affirm that you act from good motives but deny that the act itself is virtuous. I develop this strategy at length later (sec. 2.4).

The case against thinking that good motives + moralized subjective rightness suffice for virtuous action—which I will call ‘the moralized sufficiency view’—is more complicated. Two general reasons for preferring VA to the moralized sufficiency view emerge below, when we consider the relation between virtuous action and benefit. But there are also, I think, counterexamples to the moralized sufficiency view. While they are likely to be somewhat controversial, I will get them on the table here, with the intent that they function as part of the cumulative case for rejecting the moralized sufficiency view.

Consider again your pushing the man into the path of the trolley. Because many think the moral facts entail that it is impermissible to kill one to save two, many will think that pushing the fat man is (moralized) subjectively wrong. In that case the moralized sufficiency view does not entail that killing the man is virtuous, and our sense that it isn’t doesn’t count against the view.

Suppose, however, that we adjust the case. Presumably the moral facts are such that there is some number of lives saved that would justify the killing of one person. Suppose that the number is 100. (If you do not think this is enough, feel free to raise the stakes.) In the new case there is a giant train of the future (on a massively wide track), big enough to kill scores of people, heading for 100 innocent bystanders. There is the universe’s fattest man, as well as a humongous barrel, on the footbridge. As before, you believe that the barrel is not big enough to stop the train, but that the universe’s fattest man is certainly big enough. You activate the trap door under the man, instead of the trap door under the barrel, when in fact the barrel would have stopped the train.
If morality deems it permissible (and/or obligatory) to kill one to save 100, and given your beliefs about the ability of the fat man, and the inability of the barrel, to stop the train, the (moralized) subjectively right act in this case is to kill the universe’s fattest man. Suppose again we fill in the case so that your motives are ATC good: you are motivated by concern for the 100, and so on. Then pushing and killing the man will be a well-motivated (moralized) subjectively right act, and the view in question entails that it is virtuous. But I find this implausible. You have killed an innocent person, when saving the 100 did not require you to kill anyone at all. How is killing the man under these circumstances a really excellent act?

We can press this intuition further. Suppose we hold everything about this example the same except for the fact that, rather than stopping the train, the universe’s fattest man, having been dropped by you into the path of the train, ricochets off of it, dying in the process. The train speeds into the 100, killing them. Finally, imagine again that the barrel would have stopped the train. Here as before your act will be (moralized) subjectively right. The fact that the relevant empirical beliefs (i.e., that the fat man will stop the train and the barrel will not) are false does nothing to change the fact that these actual beliefs, in conjunction with the moral fact that it is permissible to kill one to save 100, supports pushing the man. So if the moralized sufficiency view is correct, killing the fat man should be virtuous. But here it seems even clearer that killing the fat man is not virtuous. You have killed an innocent person, and let 100 die, when you could have simply dropped the barrel.

Some might not share this intuition, and indeed might have the opposite sense that pushing the fat man is virtuous. But again I think this is likely to be explained by an assumption that good motives alone suffice for virtuous action, as per P-R, rather than by facts about (moralized) subjective rightness. My response to P-R’s treatment of cases emerges below, along with the rest of the cumulative case against the moralized sufficiency view. For the remainder of this section I want to consider whether moralized or non-moralized subjective rightness supplies a necessary condition on virtuous action.

2.1.2. *Is subjective rightness necessary?*

There is a second general obstacle to determining whether facts about subjective rightness place constraints on virtuous action. The distinction between virtuous motivation (which VA requires
for virtuous action) and subjective rightness is not always clear. Acts that are virtuous on VA will often as a matter of fact be subjectively right, since in many cases doing the subjectively wrong act will involve less than admirable motivation. The important question for us is whether being all-things-considered well-motivated entails doing what’s subjectively right, or, conversely, whether doing what’s subjectively wrong entails failing to be ATC well-motivated. If so, subjective rightness will be necessary, on VA (and P-R), for virtuous action.

In what follows I take it to be an open question whether we can do subjectively wrong acts from all-things-considered good motives. Some, however, may find this implausible or even incoherent on its face. It would, for example, be natural for someone who holds that doing the subjectively wrong act suffices for blameworthiness to be skeptical that we can be ATC well-motivated while doing the subjectively wrong act. How could we be both all-things-considered well-motivated and blameworthy? I will address this line of thought below.

We should see that a certain type of case would show that neither moralized nor non-moralized subjective rightness supplies a necessary condition on virtuous action. The type of case is one where someone with true moral beliefs performs an act that is virtuous but subjectively wrong. The reason that such a case would suffice to show that neither view supplies a necessary condition is that the two views imply different conclusions about which act is subjectively right only in cases where someone holds false moral beliefs. In these instances the first view makes the subjectively right act partly a function of the mistaken moral beliefs, whereas the second does not, making it instead partly a function of the correct moral principles—in which case each view returns a different conclusion about the subjectively right act. When someone holds the correct moral beliefs, this difference doesn’t apply; the two views return the same conclusion about what’s subjectively right. So if there are times when someone with true moral beliefs performs a subjectively wrong but nonetheless virtuous act, this will show that neither moralized nor non-moralized subjective rightness supplies a necessary condition on virtuous action.

4 For example, some early writers on subjective rightness (e.g., Sidgwick) identified the subjectively right act with the act the agent believes to be right. On this view doing the subjectively wrong act will often disqualify one’s act as virtuous, since acting against one’s judgment about what to do often involves a failure of motivation. While this particular view of subjective rightness is now widely rejected, it helps us to see that whether subjective rightness is necessary for virtuous action depends on whether it is possible to be well-motivated while doing what’s subjectively wrong.
Are there such cases? It seems to me that there are, and that their existence is made possible by the fact that the ‘supports’ relation—the relation that each view of subjective rightness takes to hold between the body of relevant beliefs and the subjectively right act—is objective, holding independently of whether or not the agent believes that it does. If so, we may sometimes be wrong about which act is best supported by our beliefs. In these cases we can be well-motivated in performing subjectively wrong but objectively right acts, which intuitively strike us as virtuous.  

Assuming that we can be wrong about which act our beliefs support, how might this actually happen? It might happen in exotic cases where the content of our beliefs are opaque to us, or held subconsciously. But it is more likely to happen in cases where deliberative glitches prevent us from moving from our beliefs to the right conclusion about which act they support, and so what to do. It seems to me that in some such cases we may be all-things-considered well-motivated. Being well-motivated generally helps to reduce deliberative errors, since it ensures that we will not be negligent or irresponsible in our deliberating. But it seems implausible that being well-motivated will completely insulate one from error. If the choice is made under uncertainty or pressure, and the body of beliefs is large and complex enough, even well-motivated people may err when trying to see which acts their beliefs support.

---

5 While the idea that we can be mistaken about which acts our beliefs support is a substantive assumption, it is needed if these views of subjective rightness are to avoid collapsing into the simpler and implausible view that the subjectively right act is whatever act we believe to be right. If we cannot be mistaken about which act our beliefs support, the act we think we should do will almost always be subjectively right, since the act we think we should do is almost always the one we take to be supported by how things appear to us, or our beliefs about the situation. We should also assume that the relevant body of beliefs does not include beliefs about the rules of deliberative reasoning: about how to move from our beliefs to conclusions about what to do. If they did, it would again be difficult for us to be wrong about which act is supported by our beliefs, at least insofar as we follow the rules we endorse, since our beliefs about the deliberative rules would ensure that whatever conclusion we arrive at is, in light of our beliefs about the rules, supported by our beliefs. I will thus assume that the relevant body of beliefs, those that determine the support relation, include neither our ATC judgments about what to do nor our beliefs about how to move from our beliefs toward conclusions about what to do.
Consider, for example, the case of the benevolent policymaker. Jan is in charge of policy decisions for a large aid organization. He must often make difficult decisions, involving several moral and empirical factors. The empirical factors involve considerations such as who a particular policy is likely affect, and how it is likely to affect them. The moral factors involve the values implicated in the policy: how many people it is likely to benefit, and to what extent; how many people it is likely to harm, and to what extent; how the policy stands to affect the rights and autonomy of others; how it stands to affect historically oppressed groups; how it stands to affect the health of ecosystems; and so on. Jan takes his job seriously, and strives to make good decisions. He is genuinely concerned for the well-being and rights of others, and desires policies that are compassionate and just. To this end Jan typically does not decide on policies until he has arrived at determinate beliefs about how a policy is likely to influence different segments of the organization and society at-large. But Jan often has difficulty moving from these beliefs to conclusions about what to do.

One reason for this is that moving from Jan’s beliefs to conclusions about what to do often requires Jan to engage in probabilistic and/or statistical reasoning. And Jan, while he does not recognize it, and like many others, is not very good at probabilistic or statistical reasoning. For example, Jan works with an intuitive notion of probabilities that is often responsible for his making the following type of mistake: ‘If I pass this policy, there is a 70% chance that \( p \) will obtain; if \( p \) does obtain, there is a 40% chance that \( q \) will obtain; therefore, there is a 55% chance that \( q \) will obtain. Since if I pass this policy it is more likely than not that \( q \) will obtain, and since \( q \) is the desirable end of the policy, I should pass the policy.” Of course, there is not a 55% chance that \( q \) will obtain; there is only a 28% chance that \( q \) will obtain. Rather than multiplying probabilities, Jan is taking their average. In cases like this Jan arrives at false conclusions about what act he should do, given what he believes.

Recall that we are looking for a deliberative mistake that can explain a failure to move from one’s beliefs to the right conclusion about what to do (in light of those beliefs), while at the same time leaving the ATC goodness of motives intact. The glitch in probabilistic reasoning is not essential to the case. We could, if we desired, replace Jan’s poor grasp of probabilistic reasoning with any number of common error-producing deliberative heuristics. Perhaps Jan is subject to
framing effects. Jan is more likely to conclude that he should not support a policy if his beliefs represent the policy’s chances of success under a negative frame (‘There is 30% chance that this policy will fail’) than under an equivalent positive frame (‘There is a 70% chance that this policy will succeed’). In cases like this Jan is influenced by the negative evaluative valence of the frame to make ill-supported inferences about what to do. Or perhaps Jan uses an availability heuristic that causes him to overestimate probabilities based on how easy it is for him to associate the content of relevant beliefs with other salient events in his life. Or maybe Jan is subject to a representativeness bias, one that tempts him to rely on the law of small numbers and neglect denominators, thereby leading him to overestimate the likelihood of particular events. So long as the deliberative error is common and relatively subtle, and does not involve mistakes that are blatantly obvious, it stands a chance of leaving the ATC goodness of Jan’s motives intact while explaining his failure to do what he (subjectively) ought.6

Imagine now a specific case where Jan goes astray in one of these ways. Jan must decide on some policy. After a lengthy time spent trying to discern the moral and empirical facts, Jan arrives at a body of beliefs containing all true moral beliefs but (let us suppose) one or more false empirical beliefs. Jan is not blameworthy or negligent in arriving at the false empirical belief/s; rather, the empirical facts in this case are inherently difficult to ascertain with certainty (such as how much progress on aid projects the organization will make in the coming year; how much the organization’s insurance premiums will rise/fall as a result of taking on some project; etc.). Suppose that Jan’s beliefs support not approving the policy: given what Jan believes, declining to implement the policy is the thing to do. In that case both views of subjective rightness say that Jan ought not pass the policy. Based on the sort of deliberative flaw sketched above, however, Jan mistakenly concludes that he should approve the policy, and does so. Moreover, suppose he does so because he desires to pass a compassionate and just policy, and thinks that this policy fits the bill. The pressing question is whether Jan can be all-things-considered well-motivated when failures of the sort just discussed—common and relatively subtle deliberative errors—keep him from seeing what he should (subjectively) do.

6 See Kahneman 2010 for these heuristics, and more.
The answer depends on how high we set the bar for ATC good motivation. While it’s possible to insist that Jan is not well-motivated, since he makes a deliberative mistake, this seems too rigid. For one, Jan acts from an intrinsically good desire—the desire to pass a compassionate and just policy. In addition, Jan may well have done everything we can reasonably expect of him. We can imagine that, in addition to being blameless for holding false empirical beliefs/s, Jan is also blameless for his poor grasp of probabilities (neglecting base rates, etc.). Perhaps nearly everyone in Jan’s society is subject to the glitch that leads him astray. Indeed given the extent to which error-producing heuristics appear to pervade human choice—even amongst those who should know better—to insist that their presence disqualifies Jan’s motives as ATC good threatens to make virtuous motivation a very rare thing. If ATC good motivation is incompatible with common deliberative errors, there may be hardly any virtuous motivation in the world. This seems false. But even if it is true, the sort of deliberative mistake made by Jan does not seem capable of explaining it.

We can see this even more clearly by considering situations where someone is pressed to make a high-stakes choice under time constraints. It is simply too much to ask of people in these situations that they remain impervious to all deliberative error in order to be ATC well-motivated. In high-pressure, time-sensitive situations even well-motivated agents may miscalculate probabilities, neglect base rates, overlook relevant beliefs, and so on.

Imagine that Jan gets word after lunch that his organization has been offered the opportunity to partner with a controversial group that works to deliver aid to poverty-stricken areas. The group is controversial because they have a reputation for using hard-nosed tactics, when necessary, to deliver aid to those who need it most. Jan’s boss tells Jan that they have four hours to decide whether to take the opportunity or pass it up, and delivers a stack of documents with information relevant to the choice. The documents contain information on the controversial group’s history: how much aid they succeed at delivering and to whom; what sort of hard-nosed tactics they purportedly engage in, and under what conditions; and so on. It also contains public opinion polls on the controversial group, and forecasts of the public perception of a possible partnership.
Jan takes all of this information as reliable, and sets out to see whether or not it supports partnering with the group, on-balance. Jan wants to make a good choice, and realizes that the stakes are relatively high. If he decides not to partner with the group, his organization may miss out on the opportunity to deliver aid to some very badly off people. If on the other hand he decides to partner with them, he risks negative public perception and the possibility that some important and powerful people in the aid community will regard his organization coolly, thereby limiting their ability to do future good. Unfortunately, there is a lot of data, and it is not easy to see which option it supports. As the deadline approaches and Jan’s blood pressure begins to rise, and he begins to make deliberative mistakes, unconsciously employing heuristics that lead him astray. While Jan’s beliefs about the different options support concluding that the policy is too risky, he mistakenly concludes that the risks are relatively low, and does the subjectively wrong thing. But insofar as Jan sincerely desires to make a good choice, one that is based on evidence and sufficiently responsive to the values at stake, he seems able to be ATC well-motivated. The fact that Jan is not impervious to pressure, and that this is manifested in common and subtle deliberative mistakes, should not disqualify his motives as ATC good.

The cases I have appealed to may seem contrived. Indeed they are, since this is the clearest way to make the point. But it is worth pointing out that the basic phenomenon in question—where deliberative glitches keep us from seeing which act is really supported by our beliefs—is not confined to complicated cases like Jan’s. Since the heuristics in question always work on one’s beliefs, or how it seems to us about some situation, they will often cause error by eliciting faulty inferences from a body of beliefs to a conclusion that is not really supported by them. Whenever this occurs in a deliberative chain of reasoning—a reasoning process aimed at discerning what to do (or actually doing it)—we can have agents who reach the wrong conclusions about which act their beliefs support. In at least some of these cases, I claim, the person’s motives may yet be ATC good.

This is not to say that deliberative mistakes are irrelevant for whether some episode of motivation is ATC good. For one, there may be times when people who reason from their beliefs to wrong conclusions about what to do are negligent, and really should have done better.
If we can reasonably expect more of someone, the error may disqualify their motives as ATC good.

Moreover, it is compatible with everything I have said so far that being subject to deliberative glitches detracts from the goodness of motives and is pro tanto bad. The key question is whether deliberative mistakes always suffice to disqualify motivation from being all-things-considered good. While the sort of deliberative mistakes I’ve appealed to plausibly detract from the goodness of motives, it is far less plausible that they always suffice to disqualify motives as ATC good. If Jan desires to pass a compassionate and just policy, believes that this policy is just, and has done everything we can reasonably expect him to do, we may very well want to conclude that, even if Jan’s motives aren’t perfect, they are all-things-considered good.

If Jan is all-things-considered well-motivated in these cases, we have a well-motivated but subjectively wrong act. Defenders of P-R will in this case recognize Jan’s act as virtuous, since it is well-motivated, and should conclude that neither notion of subjective rightness is necessary for virtuous action. Whether Jan’s act is virtuous on VA, rather than just well-motivated, depends on whether Jan’s act turns out objectively right. If Jan’s beliefs were all true, they would presumably support the objectively right act. In that case it wouldn’t be possible both for Jan to do the act that he does—the one not supported by his beliefs—and for that act to turn out objectively right. But if (as in the original example above) Jan has one or more false empirical beliefs, there is nothing to prevent his act from being objectively right, even if it is not supported by his beliefs. Since there is nothing to guarantee that all of Jan’s empirical beliefs will be true, people in his situation may wind up doing the subjectively wrong but objectively right act. If Jan’s act does turn out objectively right, VA also deems it virtuous. If so, neither view of subjective rightness states a necessary condition on virtuous action, for either P-R or VA.

---

7 Note that a close analogue to Jan’s case shows that subjective rightness is not sufficient for virtuous action, either. All we need do is switch the case so that Jan is poorly motivated and fails to do the act supported by his beliefs. Perhaps, for example, Jan desires to pass the policy he (mistakenly) takes to be right, not because it is right or compassionate or just, but because he knows that this particular policy will anger his ex-partner, whom he wants to suffer.
I’ve been arguing that certain deliberative mistakes responsible for agents doing the subjectively wrong thing are compatible with virtuous motivation, and so that neither moralized nor non-moralized subjective rightness states a strict necessary condition on virtuous action. It might be thought that all of this glosses over a clear problem with denying that subjective rightness is necessary for virtuous action. The real difficulty, it might be thought, arises from cases where it is obvious, at the time at which someone acts, that they should not do some act; where they do the act anyways; and where the act winds up, by blind luck, objectively right.

Suppose that Dave is drowning off the dock. Sue is not a strong swimmer, and Dave is twice her size. Her beliefs about herself and the situation clearly support not jumping in to try to pull Dave out herself. But Sue is overcome with concern for Dave, and jumps in after him. By a massive stroke of luck Sue pulls Dave to safety. Sue has therefore done the objectively right thing. But if subjective rightness isn’t necessary for virtuous action, how can we conclude that Sue doesn’t act virtuously?

It should be clear how the approach I prefer will respond. If Sue does fail to perform a virtuous act, this can be explained in terms of a failure of motivation, rather than as a failure of subjective rightness per se. On one way of filling in this case Sue believes, all-things-considered, that she should try to save Dave. Here Sue is like Jan: while her beliefs in fact support not jumping in, she mistakenly concludes that she should. But if it is really that obvious that Sue should not jump in, Sue must be careless and negligent to conclude that she should. Someone who is so careless and negligent is not ATC well-motivated, even if the motivating desire (e.g., a concern for Dave’s welfare) is intrinsically good. In that case we do not need to appeal to subjective rightness, over-and-above the failure of motivation, to explain why Sue’s act isn’t virtuous. If on the other hand it is not so obvious which act Sue’s beliefs support, it is no longer clear her act is not virtuous.

On a second way of specifying the case Sue believes that she should not jump in, but does so anyways. Here again we can explain any sense we have that Sue’s act is not virtuous by appeal to motivational deficiency. Sue acts against what she believes she should do. She is indifferent.
to what she believes to be right. Such weakness is often considered a paradigm case of irrationality. As such, it is natural to find it inconsistent with ATC good motivation.  

I’ve argued that, on both moralized and non-moralized subjective rightness, it is possible to be ATC well-motivated while doing subjectively wrong but objectively right acts. If so, neither view of subjective rightness provides a necessary condition on virtuous action, as that notion is understood by either P-R or VA. While it does not follow that no specification of subjective rightness provides a necessary condition on virtuous action, the arguments of this section do provide evidence for thinking that we will be able to explain everything that needs to be explained in terms of failures of virtuous motivation, not subjective rightness per se.

So far I have motivated VA by showing how the organic unity interpretation provides an illuminating way to think about virtuous action, and have argued that we do not need a subjective rightness constraint on virtuous action. I want now to return to my earlier claim that VA implies distinctive claims about the relation between virtue and the good.

2.2. Virtue and benefit

The aim of this section is to point out two different ways in which VA differs from P-R. These differences involve the set of claims each view implies about the relation between virtuous action and the good. First, VA and P-R return different conclusions about the relation between virtuous acts and personal benefit: the way in which, and the extent to which, virtuous action increases the good of its possessor. Second, VA and P-R differ in the set of claims they imply about virtue and social benefit: the way in which, and the extent to which, virtuous actions increase the goods of communities. After laying out the ways in which the views differ, I argue that VA has an easier time than P-R explaining plausible claims about the relation between virtue and benefit.

---

8 This treatment of the second version of the case might appear controversial, since some hold that acting against one’s ATC judgment is consistent with virtuous motivation. In particular, some have argued that agents who act against mistaken beliefs about what they should do—so-called ‘inverse akratics’—are capable of being well-motivated. (See Bennett 1974; Driver 2001; Arpaly 2003). But we needn’t think that inverse akratics are incapable of being well-motivated for the current strategy to work. When acting against one’s judgment is consistent with being ATC well-motivated, the act can be virtuous; when not, not. In neither case must we appeal to facts about subjective rightness to get what we want.
2.2.1. Personal well-being

The view that virtue benefits its possessors is one of the oldest ideas in Western philosophy, and one of the most contested as well. While there is and ought to be controversy over the details of this view—e.g., over how much virtue benefits its possessor; whether it improves the subjective well-being of its possessor or is beneficial in some more objective sense; and so on—it seems to me hard to deny that virtue makes our lives better, on some relevant scale. With respect to VA, there is an obvious sense in which someone who realizes good ends from good motives will typically benefit. Someone who from good motives succeeds at respecting and benefiting others, satisfying social obligations, and so on, is likely to receive the admiration and esteem of her neighbors, as well as the opportunities that come to well-liked and dependable people. The connection VA posits between virtue and objective rightness provides a nice explanation for why virtuous people often enjoy goods of this sort.

Of course, many have held there to be a much stronger connection between virtue and personal benefit. The ancient Greeks, for example, held that being virtuous benefits us in a much more direct way: by being partly constitutive of, or perhaps even identical to, living a good human life. This claim is nowadays difficult to defend. It looks like being virtuous sometimes, perhaps often, requires us to sacrifice a measure of personal well-being. Justice, for example, may require us to give up goods that would greatly enhance our satisfaction with life; the same goes for generosity, loyalty, and the rest of the other-regarding virtues.

Many have thought that the root of this problem stems from the fact that the sense of ‘happiness’ or ‘the good life’ to which the Greeks and others historically took virtue to contribute differs markedly from contemporary views about well-being. Whether or the extent to which this historical claim is true, it seems very likely that there are simply different scales on which the goodness of lives can be measured. Some of these scales are subjective, in the sense that they give the perspective of the person whose life it is sole weight for determining whether that life is going well. Objective views, by contrast, put limits on which attitudes can increase well-being, typically claiming that pleasure and the satisfaction of desires increases well-being only when those attitudes are good in some more objective sense: only if they are intrinsically good and/or

---

9 Sumner 1995. Cf. Kraut 1979 for a nifty argument that Aristotle’s notion of eudaimonia of necessity involves a more subjective brand of happiness, which I think (so far forth) undermines the mantra that Greek eudaimonists and moderns employ wildly different notions of happiness.
aim at something of real worth. Someone who satisfies her actual desires and aims may be doing well on the subjective scale without doing well on the objective scale, since her desires and preferences may not aim at real goods. Similarly, someone who fares well on the objective scale may not be pleased or satisfied with his life, since he may not recognize or appreciate the objective goods in it.

On this way of setting things up both P-R and VA plausibly imply that virtuous acts contribute to personal well-being on the objective scale, and to do so because of the intrinsically good or admirable motivation they involve. This is attractive. All of us admire people who exhibit the motives of courage or generosity or justice, because they have these motives. The motives of Martin Luther King, Dorothy Day, Paul Farmer, and so on, are part of what make their lives good and fitting objects of admiration. Someone whose life is full of admirable and intrinsically good motivation is so far forth leading a good life, in at least one important sense. P-R and VA both nicely capture this thought.

But whereas P-R claims that the benefits of virtuous action are exhausted by the good motives from which virtuous acts are done, VA goes further, in two ways. First, proponents of VA can also claim that virtuous action contributes to well-being via permissible success. Second, VA can claim that virtuous action contributes to well-being via the extra value that exists in the combination of admirable motivation + permissible success. It seems to me that both of these claims are plausible, and that this gives us reason to prefer VA.

First, it seems plausible that permissible success increases objective well-being. We prefer a life full of permissible success to a life full of failure. The value of success is perhaps clearest in cases where the success is extremely valuable. The success of Mandela or Martin Luther King clearly seem to make their lives better; we admire them partly because they succeeded, and more than we would if they had failed. But we can also see the value of permissible success in more mundane cases, such as those discussed earlier: we prefer to succeed at making our partner happy, to keep promises, to be a reliable colleague or spouse or parent. Looking back over our lives we are pleased at the successes and saddened by the failures, even if the failures are not our fault. While I think only permissible success increases the value of lives—we could not justifiably look black with pleasure or approval at a life full of successful harming or lying or cheating—it does not seem strictly necessary for success to be well-motivated in order to
contribute individual benefits. Other things equal, a life full of neutrally or even badly motivated permissible success, while of course less good than a life full of well-motivated permissible success, is still better than a life full of badly-motivated failure.\(^\text{10}\)

If permissible success does make our lives better, VA can claim that virtuous action contributes to well-being not merely by way of virtuous motives, but also via the permissible success it involves. This too is attractive. As I said earlier, many have held that virtue benefits its possessor. Indeed it is tempting to think that this holds as a conceptual truth, given plausible metaethical assumptions about virtue, namely, that virtue-properties make excellent that in which they inhere. To the extent that virtuous acts are part of an individual lives, those lives should be made better by their presence. While many have thought this ‘making better’ is accomplished by the presence of virtuous motives, VA shows how acting virtuously can make our lives better in a second way: by ensuring permissible success. In so doing VA strengthens the connection between virtue and benefit in an independently plausible way.

This is not VA’s only attractive implication for personal well-being. Given that virtuous action contributes to well-being by way of both admirable motives and permissible success, virtuous acts will, on VA, also contribute to well-being via the value that resides in their combination. Consider someone whose well-motivated attempts to realize permissible ends fail, but who later realizes the same ends from neutral motives. On Tuesday Alice plans to make her partner Tim happy, by writing and giving him a poem. Alice’s motives on this occasion are admirable: her poem-writing is motivated by an occurrent love for Tim, and a desire to make him happy for its own sake. Alas Alice does not succeed in her aim: the poem she writes is swept out of her hand by the wind as she walks up to Tim’s house. On Thursday Alice again decides to make Tim happy by writing him a poem. However, this time Alice is motivated by a neutral motive: she desires to give Tim a poem from the motive of self-interest, in the belief that doing so will reflect

\(^{10}\) The ‘other things equal’ caveat is important. We might be tempted to think that a life full of permissible success is better than a life full of failure because a life full of success will have more of the instrumental goods of success, and in particular more subjective satisfaction, than a life full of failure. But I don’t think this explains our preference for a life full of permissible success. We can see this if we stipulate that the life full of failure is equal in subjective satisfaction to a life full of permissible success. It still seems we should prefer a life of permissible success. It might also be that an attempt’s being badly-motivated decreases the value of a permissible success; this would, I think, be a natural view for defenders of VA to hold.
favorably on her. (If you do not think the motive of self-interest is neutral under these conditions, feel free to fill in the case with a motive that is.) This time Alice succeeds; she gives Tim the poem, which makes him happy.

Compare now this first scenario with a second scenario, one where Alice’s initial well-motivated Tuesday attempt succeeds. The organic unity interpretation predicts that, other things equal, there is more value in the second scenario than in the first, since the second scenario does, and the first does not, contain the extra value that comes from combining admirable motives + permissible success. Our question is whether this extra value contributes to personal well-being.

It seems that it does. Imagine a life full of disjointed episodes of the sort we find in the first scenario (well-motivated failures to do good deeds, combined with neutrally motivated later success at doing the same act), and a life full of well-motivated successes. Even if the same ends and motives are realized in both cases, the second life seems better (other things equal). The second life contains a sort of satisfying unity the first life lacks. We care not only about the number of permissible successes and good motives in our lives, but about how they are combined. Other things equal, we prefer our well-motivated attempts to succeed, and our successes to be well-motivated, even if no good motive or permissible end is lost. The second life seems better than the first in these regards. VA can explain this by claiming that virtuous acts contribute to personal well-being not only by ensuring permissible success, but also via the value that inheres in the combination of admirable motives + permissible success. Here again VA strengthens the connection between virtue and individual benefit in an independently attractive way.

While P-R can agree with VA that virtuous action makes our lives better via virtuous motivation, it denies that virtuous action contributes to well-being via permissible success. It isn’t immediately obvious, however, whether or to what extent this constitutes an important difference between VA and P-R. For there is nothing to stop proponents of P-R from recognizing that permissible success matters for well-being. A proponent of P-R who does so will, like

---

11 On P-R acts that are virtuous may be objectively right, in which case they will contribute to well-being. But this is contingent. In these cases what explains the gains in well-being—the success—is not essentially tied to the act’s virtue.
proponents of VA, be able to capture all of the benefits of permissible success for personal well-being. P-R will not, of course, explain these benefits under the concept of virtuous action.

Nonetheless, insofar as the preceding arguments are plausible, VA is still likely to have an advantage here. This is because, even if P-R makes permissible success a constituent of well-being, P-R struggles to capture the full relevance of well-motivated permissible success for well-being. To capture the full relevance of these acts for well-being, P-R must do more than make permissible success a constituent of well-being. P-R must, like VA, say that the combination of admirable motives + permissible success per se makes our lives better. If the additional value in these wholes contributes to personal well-being, P-R must be able to recognize this value to capture everything VA does about personal benefit.

As I argued earlier (sec. 1.5), however, P-R has difficulty accommodating the thought that admirably motivated permissible success constitutes an organic whole. Proponents of P-R cannot claim this without either making the virtue of acts depend partly on permissible success, or conceding that the value of motives is not fully reflected in the virtue of acts. Since P-R cannot take the first horn without collapsing into VA, and since the second horn is independently unattractive, VA has an easier time explaining the individual benefits of admirable motives + permissible success.

I’ve argued that, in addition to the benefits of admirable motivation, virtuous action increases personal well-being by way of the permissible success it involves, as well as by way of the value found in wholes consisting of good motives + permissible success. If we find these claims plausible, and insofar as P-R struggles to capture the idea that well-motivated permissible success form organic wholes, the arguments of this section give us reason to prefer VA to P-R.

The arguments of this section can also form part of the cumulative case (cf. sec. 2.1) for rejecting the moralized sufficiency view: the idea that good motivation + moralized subjective rightness suffices for virtuous action. Since subjective rightness does not guarantee objective rightness, the benefits of permissible success, as well as the value of well-motivated permissible success as a whole, will not accrue in the life of someone who acts virtuously. They may accrue—but this will be entirely contingent on whether our subjectively right acts are also objectively right. This
gives those of us who think that there is a tighter connection between virtuous action and individual benefit reason to prefer VA to the moralized sufficiency approach.

2.2.2. Social goods

In addition to virtue’s personal benefits, many hold that the virtues facilitate social interaction, helping to secure or increase important social goods. This seems plausible on its face. A community of caring and generous people will be more harmonious—a better place to live—than a community of stingy scoundrels. A community is better if its members are trustworthy and considerate and just, rather than selfish jerks. How exactly does virtuous action promote social goods?

One way is via the admirable motives it involves. If I know that you have goodwill toward me, or care about keeping your promises, this will structure our relationship in positive ways. I may, for one, feel pleased and secure by the knowledge that you care about what actually matters. I am also more likely to act in mutually beneficial ways, to co-operate with you and others, when I believe that my efforts will be reciprocated, and that the people around me are not free-riders. Many co-operative ventures require trust. Knowledge of each other’s good motives and intent can help keep co-operative ventures from devolving into pernicious collective action problems.

VA and P-R can both affirm the social goods that admirable motivation makes possible. But VA can again go further, by ensuring permissible success. There are two senses in which permissible success might contribute to social goods. First, permissible success might increase social utility, by promoting well-being. On any plausible brand of consequentialism, for example, permissible success will often involve promoting human well-being. This might happen directly or indirectly. It happens directly when permissible success consists in (say) making someone happy, or relieving their distress. It happens indirectly when permissible success contributes to the long-term well-being of those in a community. Even if we have sufficient reason apart from human well-being to preserve a tract of land—say, because it

12 This claim is perhaps most often made by consequentialists who identify virtues as features of persons that promote utility in populations. See Hume 1739 (bk. 3, part 3, sec. 1) and 1777 (ch. 8); Bentham 1780 (esp. secs. 10.33 and 11.2); Driver 2001; Bradley 2005. But consequentialists are not the only ones who extol the social benefits of virtue. See Foot 1978, Thomson 1997, and Zagzebski 1996 for examples of non-consequentialists who tie virtue and social benefit closely together.
contains rare and endangered species, or complex well-functioning ecosystems—doing so may yield long-term benefits to the members of the community: opportunities of education, knowledge, and virtue, in addition to happiness. Any permissible success that involves benefiting others in these ways will raise social utility in this first sense.

There is also a wider sense in which permissible success can benefit communities, one that is relevant if deontological facts determine right and wrong. The more the moral facts are deontological, the less likely it may be that permissible success promotes well-being. Imagine a deontological theory on which considerations of justice usually take precedence over promoting well-being. On this view permissible success will fail to promote well-being whenever respecting justice, rather than promoting well-being, is required.

Nonetheless, there is still a clear sense in which permissible success contributes to the good of communities. It does so simply by making it just. A community where people follow the dictates of justice is, in one clearly recognizable sense, better than a community where people do not, even if the levels of utility in the just community are relatively low.

Of course, utilitarians are likely to bristle at this notion of what counts as ‘social goods’ or ‘a good community.’ My claim here is simply that there is some recognizable sense in which acting justly and respecting rights (and so on) make a community a better place, even if these are not correlated with increases in human welfare. This is fully compatible with holding that well-being matters for the good of communities. Just how we should weigh justice or rights against utility in communities is of course highly disputed, and I cannot pursue it here. For now my claim is just the relatively weak one that a community full of people who act fairly and justly is (pro tanto and ceteris paribus) better than a community of people who do not, even if this is only one factor among others on which the goodness of communities depends. If so, permissible success will be so far forth good for communities, even if the moral facts are deontological.

While P-R can guarantee whatever social goods come with admirable motivation, it cannot guarantee the goods of permissible success. Since no amount or degree of good motivation can ensure that promises are kept, wrongs repaired, suffering alleviated, and the good of others promoted, virtuous action will not, on P-R, ensure the social goods associated with these results. Moreover, the fact that everyone in a community is well-motivated says nothing about what ends
they are well-motivated to pursue. If it is possible to intend and/or do the wrong thing from
good motives, good motivation might wind up resulting in significant harms to a community.
This could happen in cases where agents are non-culpably mistaken about which ends are good
and which bad, and pursue objectively bad ends, non-culpably thinking them to be good. If this
phenomenon were widespread enough, it will be possible for a community of people who act
virtuously on P-R virtue to be a positively bad place to live.

This said, such scenarios seem pretty unlikely, at least in worlds resembling ours. In general,
while admirable motivation doesn’t guarantee right acts, or the goods of community, it does
make them more likely. The extent to which P-R’s inability to guarantee the same social goods
as VA licenses us to prefer VA depends on how tight we take the connection between virtue and
social goods to be. In this regard a defender of P-R might justifiably point out that, with enough
bad luck or (non-culpably) false beliefs, any good ability or power can fail to produce social
goods. If some of these failures have nothing to do with virtue, the fact that VA guarantees
social utility does not give us reason to prefer VA to P-R.

What we think here depends on how much disconnect we are willing to tolerate between virtuous
action and the goods of community. If we are willing to tolerate large discrepancies between
virtuous action and social goods—if we are unperturbed by the thought that a community could
be full of people acting virtuously but nonetheless fail to be a good, and even be a bad, place to
live—we will see here no reason to prefer VA to P-R. But it seems to me that there is at least
some dissonance in this thought. This dissonance is, again, predicted and explained by plausible
metaethical views about the functional role of virtue-properties in making things good. If one of
the things that virtue-properties make good are social communities of people, no less than
individuals themselves, groups of people who consistently act virtuously but comprise lackluster
or even bad communities is an intuitively jarring thought.

At the same time, because the disconnect between virtuous action and social goods will, on P-R,
at least sometimes be due to bad luck, those committed to thinking that virtue is immune to luck
are unlikely to find the disconnect jarring. Such people must be willing to accept that virtue
sometimes fails to make communities good in the ways sketched above. But they may be willing
to do so, given their other commitments. While for me the dissonance in this thought raises my
credence that acting virtuously requires good luck, I do not know what more to say to someone
who lacks this intuition, or how to press the argument further without begging any questions against those who deny that virtue is luck-dependent. Those who deny that virtuous action can be disrupted by luck will be able to resist concluding that the arguments of this section support VA over P-R. Those more inclined to see virtue as luck-dependent will, on the other hand, see reason to prefer VA.

Even if VA explains the social goods of virtuous action better than P-R, it might be objected that the most straightforward way to explain virtue’s social utility is not by appeal to objectively right action, but simply by appeal to the good. This seems true. As I argue in the next chapter, however, the very features that allow standard consequentialist theories of virtue to provide a stronger explanation of social utility give rise to serious difficulties, not the least of which is their inability to capture an adequate connection between virtue and the right. For now I can rest content with the claim that VA provides a very good explanation of the social goods of virtue, and in particular a better explanation than P-R. If the arguments in chapter three work, we’ll have strong independent reason to reject pure consequentialist theories, and so reason to accept VA as the best working explanation of virtue’s social goods.

Finally, like the considerations about personal well-being in the last section, the arguments of this section can form part of the cumulative case for rejecting the moralized sufficiency view: the claim that good motives + moralized subjective rightness suffice for virtuous action. Unlike objectively right acts, doing subjectively right acts does not ensure communities with high levels of well-being. Nor does it ensure that rights are respected, or that communities are fair and just. Again, subjectively right acts may accomplish these things, but this will turn on what sort of empirical beliefs members of a community have; whether these beliefs support acts that make communities better; and (if so) whether community-members succeed at realizing these ends. These are, of course, contingent and uncertain matters. If we think the connection between virtuous action and good communities is stronger than this, we have reason to prefer VA to the moralized sufficiency view.

2.4. Cases

I’ve been arguing that VA implies plausible claims about the relation between virtuous action and benefit, and that this is due to the place it affords objective rightness in the explanation of
virtuous acts. It may yet turn out, however, that this very feature of VA—its insistence that acts be objectively right—renders it ill-equipped to capture considered judgments about the virtue-status of particular acts. Because VA and P-R disagree over whether an act must be right to be virtuous, they return different conclusions about the virtue-status of acts. Our question is which (if either) theory returns better conclusions: verdicts that more closely match our considered judgments.\(^{13}\)

Given that the main nub of disagreement between VA and P-R is over whether well-motivated acts must be right to be virtuous, the cases we consider should isolate this difference. We can do this by focusing on well-motivated actions that fail to be objectively right. Within this class there are at least four main types of test cases: (i) those where agents are virtuously motivated but do the objectively wrong thing because of justified but false empirical beliefs; (ii) those where agents are virtuously motivated but do the objectively wrong thing because of unjustified false empirical beliefs; (iii) those where agents are virtuously motivated but do the wrong thing because of justified but false moral beliefs; and (iv) those where agents are virtuously motivated but do the wrong thing because of unjustified false moral beliefs.

It will ultimately be controversial whether the agents in (ii) and (iv)—people with unjustified false beliefs, either factual or moral—can be genuinely virtuously motivated. It’s also not clear whether there are genuine instances of (iii), since it’s controversial whether one can be justified in holding bad moral beliefs, in such a way that one’s motives remain good. For this reason, and because cases of type (i) are at any rate likely to serve proponents of P-R best in the debate with VA, I’ll restrict my attention to cases of well-motivated individuals who do the wrong thing from justified but false factual beliefs. Our main question is whether there is reason to prefer a P-R or VA analysis of them. These cases will also let us continue to assess the possibility that good motives + (moralized) subjective rightness suffice for virtuous action.

Here are two cases that fall under (i).

---

\(^{13}\) While the earlier cases in section 1.3.4 (Saving One) were designed to test our intuitions about whether certain acts intuitively form organic wholes, these will be designed to test more directly whether certain acts are intuitively virtuous.
Percy

Percy has a strong desire to help others, and believes that the best way for him to help is to donate money to organizations that focus on sustainable development in poor nations. Percy’s generally reliable colleague George recommends one organization as particularly effective. After a good bit of research—all the research we could reasonably expect of him—Percy concludes that the outfit is honest and respectable, and that he will donate money to them. Unfortunately, Percy (and George) are wrong; the people to whom they give their money are very skilled at deception, and not on the up-and-up. The outfit takes Percy’s money and secretly funnels it to the de facto military leaders of the nation, who use it to promote causes that harm the worst off. Percy has acted out of concern for the well-being of others, but has wound up harming, rather than helping, those in need. Is his action generous?

Richard

Richard, a trial court judge, is trying someone charged with grand larceny. He is motivated by an admirable desire—to be fair, and to see justice be done. Suppose that in reality the individual being tried is innocent, but that the preponderance of available evidence supports the conclusion that he is guilty. On this basis Richard convicts the defendant. Is Richard’s act just?

In each of these cases P-R says that the act in question is virtuous (generous and just, respectively), because it is well-motivated. Moreover, it is easy to stipulate, for each of these cases, that the acts are (moralized) subjectively right. Suppose we do so: the acts of Percy and Richard are sufficiently supported by their empirical beliefs, in conjunction with the moral facts. These cases would then isolate the relevant difference between both VA and P-R, and between VA and the moralized sufficiency view. VA and the moralized sufficiency view entail that the acts in question are virtuous. VA, on the other hand, entails that they are not. Which entailment is more plausible?

It’s important to keep in mind that the question is whether the acts in the examples are generous and just, not over whether the agents are praiseworthy or blameworthy, or well- or badly-motivated. Most everyone is likely to feel a measure of admiration for the agents in the examples, to the extent that they are well-motivated. VA has no trouble with this; it agrees that
the agents in the examples act from admirable motives, and that people such as these, individuals
moved by good motives to do what’s objectively wrong are, to the extent that they are well-
motivated, worthy of admiration.

But once we see this it looks like VA gives us a way to capture everything we really want about
the cases. VA allows us to recognize what is good about the actions in the examples—that they
are performed from generous and just and temperate motives—and lets us praise the agents for
this, without committing us to thinking that harming others could be generous, or convicting an
innocent person just. This general strategy can be deployed against other possible
counterexamples to VA. In each case VA can deal with the counterexample by distinguishing
the admirable motives from which the act is done, on the one hand, from the virtue proper of the
act, on the other, and claiming that the distinction actually makes better sense of the cases, by
making it possible for us to admire the agent while satisfying our sense that a truly generous act
(for example) must not harm, but help.

A sophisticated objector might at this point grant that there is some sense in which these acts fail
to be generous and just, but that it is not the sense relevant for virtue. Focus for now on justice.
‘Just’ and its cognates, the objection goes, can be used both to ascribe virtue-properties and in a
more purely deontic sense. When ‘just’ is used in a virtue sense it implies claims about motives
or inner states; when it is used in a deontic sense it does not. We see the deontic sense of ‘just’
when we say that some outcome or distribution of goods is just. That something is a just
outcome or just distribution doesn’t imply that certain motives, just ones, are behind the outcome
or distribution. But perhaps then our sense that Richard’s act of convicting an innocent man is
not ‘just’ is not an intuition about the virtue sense of ‘just,’ but the deontic one. This would
amount to saying that Richard’s act is virtue-just, because it is justly motivated, but not deontic-
just, because it brings about an unjust outcome, or is (more simply) wrong. Moreover, perhaps
something similar holds with the other virtues. Perhaps Percy’s act is virtue-generous, because

---

14 I owe this line of response to Tom Hurka.
generously motivated, but not deontic-generous, because of its bad results; and so on. Call this
the ‘dual-usage’ view of virtue-concepts.\textsuperscript{15}

Both the dual-usage view and VA agree that the agents in the examples are virtuously motivated,
and that the acts they perform in some sense miss the mark. Where VA parts ways with the dual-
usage theory is over whether motives are the only thing that matter for determining the virtue-
status of acts. On dual-usage theory, whether an act is ‘generous’ or ‘just’ in the deontic sense
does not determine its virtue-status; this is the point of distinguishing the virtue and deontic
senses of virtue concepts. The virtue-status of acts is determined just by the virtue-sense of the
relevant terms (by whether the act is, for example, virtue-generous or virtue-just), which depends
solely on facts about motives. Dual-usage is thus a version of P-R, claiming that motives suffice
to determine the virtue-status of acts.

We should see that, because of this, the dual-usage view will not secure any of the benefits that
come from capturing the value of wholes comprising admirable motives + permissible success
beneath the concept of virtuous action. Granted the organic unity interpretation of VA, acts that
are, in dual-usage terms, both virtue-just (for example) and deontic-just will be loci of additional
value, by forming organic wholes consisting of admirable motivation + permissible success.
Since the dual-usage theory does not require acts to fare well along both axes (to be both virtue-
just and deontic-just) to warrant virtue-terms—if it did it would collapse into VA—the dual
usage theory does not capture the value of admirable motives + permissible success beneath
virtue-concepts.

Moreover, we can ask what’s really to be gained by complicating the theory of virtue in this way.
Dual-usage tries to explain whatever sense we have that the acts in the example aren’t generous
or just, despite the fact that in each case the agent is admirably motivated. It says that Richard’s
act is just, by virtue of being justly motivated, and unjust, by virtue of failing to bring about a fair
outcome. But why should we prefer this to the seemingly simpler thought that Richard is justly

\textsuperscript{15}It may be that some virtue-concepts (such as justice) admit of dual usage, while others (such as
generosity) do not. This would limit the scope of the objection. Since I think the arguments that
follow support rejecting the dual-usage strategy for all virtue-concepts, I won’t fuss this point.
motivated but performs an unjust act? (Mutatis mutandis for Percy.) We need some reason. To insist on a dual-usage theory merely because it allows us to keep the virtue of acts solely a function of motives is, in this context, to beg the question against VA. Given the overall attractiveness of VA, and its simple strategy for handling cases of well-motivated objectively wrong acts, the dual-usage theory of virtue-terms appears under-motivated in relation to VA.

This said, some might doubt that VA’s general strategy for handling the cases—acknowledging the virtue of motives while denying the virtue of acts—gives us the intuitively right conclusions about all cases where virtuously motivated people do the objectively wrong thing. VA implies, for example, that someone who is courageously motivated to attempt to save a drowning friend, but who suffers a freak leg cramp and drowns in the process, fails to act courageously.

How damaging this is to VA depends on the strength of our sense that this person does a courageous act, rather than merely being courageously motivated. This is a fairly subtle difference. Ultimately, it may be difficult to conclude that either VA or P-R clearly outperforms the other in this and similar cases. We may have a fairly strong considered sense that the person who drowns trying to save a friend does something courageous, rather than just being courageously motivated. If so, this supports P-R over VA. But we may also have a fairly strong considered sense that convicting an innocent man is not just, and that harming those in need is not generous. If so, this supports VA over P-R. At the level of cases, it seems hard for either VA or P-R to claim a clear victory. Mutatis mutandis for VA and the moralized sufficiency view.

This said, if my earlier arguments are correct—if VA provides a compelling account of virtuous action as an organic unity, and does a better job than P-R and the moralized sufficiency view at capturing abstract relations between virtue and other moral concepts—a stand-off in the realm of cases may be better for VA than its competitors. I hope at least to have shown that VA is a viable alternative to P-R, one that deserves to be taken very seriously, and which may ultimately do a better job than P-R and subjective rightness-based views at capturing attractive ideas about the nature of virtue and its relations to other moral concepts.
Where does this leave us? Clause (ii) of VA, recall, claims that acts are virtuous only if and partly because they are right. Are we now justified to accept at least the logical claim, that acts are virtuous only if they are right? Not yet. Even if the arguments of the first two chapters warrant us to prefer VA to P-R and subjective rightness-based approaches, they do not warrant us to accept VA outright, since there may be other views of virtuous action that (like VA) entail clause (ii) but are preferable to VA, all things considered. In particular, consequentialist theories that explain virtue entirely in terms of a productive relation between virtue and the good may, for all I have said so far, both imply clause (ii) and be preferable to VA. Chapter three is devoted to this possibility.
Chapter three
Virtue and consequence

3.1. Virtue-consequentialism

In the first two chapters I began to argue for the following view of virtuous action.

Virtuous Action (VA): Acts are virtuous just in case and because they are (i) all-things-considered well-motivated and (ii) all-things-considered objectively (at least) permissible.

I argued that we have reason to prefer VA to approaches that do not imply a connection between virtue and rightness, as well as to views that make virtue depend on subjective, rather than objective, rightness. I did this by advancing an attractive interpretation of VA, according to which well-motivated objectively right acts constitute organic wholes with their own distinctive value; and by arguing that VA does a better job than both subjective rightness-based views and the Prichard-Ross approach at capturing connections between virtue and other moral concepts.

In this chapter I continue to defend VA by arguing for its superiority over another prominent conception of virtue: consequentialist approaches that make virtue solely a function of consequence.

Virtue-consequentialism differs from traditional act-consequentialism in two ways. First, act-consequentialism is a theory about deontic properties; its purpose is to tell us which acts are right and which wrong, and to explain why. By contrast, virtue-consequentialism is a theory about virtue-properties; it explains which entities are virtuous and which vicious, and why. Less obviously, virtue- and act-consequentialism can differ in their objects of assessment. While act-consequentialism takes acts as its sole evaluand, I’ll assume that virtue-consequentialism takes inner states of persons as a direct object of assessment. Some prominent virtue-consequentialisms take traits of character as their evaluand; these views say that traits of character are virtues just in case and because their consequences are good enough. It is also possible for virtue-consequentialisms to take occurrent motives as their object. What about acts?

Given the basic idea behind virtue-consequentialism—that virtue is a function of consequence—a natural thought is that acts are virtuous when and because their consequences are good enough.
Virtue-consequentialists who are also act-consequentialists will then claim both that the virtue of acts and their deontic status depends on the amount of good they produce.

For such consequentialists, however, the distinction between deontic and virtue properties is obscure. If acts are both right and virtuous when and because (for example) they maximize the good, it’s not clear that we need both deontic and virtue properties, or what work one of them does that the other doesn’t.

Consequentialists can avoid making the virtue and rightness of acts co-extensive by making virtue and rightness depend on realizing different amounts of good. Acts could be right when and because they produce $n$ good, and virtuous when and because they realize $n + 1$. But now being a virtuous act just amounts to having slightly better consequences than right acts. We don’t need the concept of virtue for that; virtue does no real work. Conversely, being a right act amounts to having slightly worse consequences than virtuous acts, and rightness does no real work.

This conclusion can be avoided by giving up either virtue- or act-consequentialism, but most true-blue consequentialists will presumably resist doing so. Alternately, virtue-consequentialists can say that the virtue of acts is at least partly derivative: acts are made virtuous at least partly by proceeding from more fundamental bearers of virtue, such as motives or character traits, whose virtue-status is explained directly in terms of the good. I’ll assume that the most plausible virtue-consequentialist accounts of virtuous action claim that acts are made virtuous at least partly by their connection to virtuous inner states (more on this below).

What is the relation between virtue-consequentialism and the arguments for VA advanced in chapter one? Importantly, some of the considerations that recommend VA over the Prichard-Ross approach do not recommend VA over virtue-consequentialism. For example, while appeal to the social benefits of virtue gives us reason to prefer VA to P-R, it seems more likely to recommend virtue-consequentialism over VA, rather than vice-versa.

Moreover, whereas the Prichard-Ross approach builds a disconnect between virtue and rightness into the view, it isn’t obvious that virtue-consequentialism cannot capture a strong connection between virtue and rightness. Here virtue-consequentialists must proceed with caution. As I understand them, virtue-consequentialisms are ambitious: they claim that the good *alone*
provides a complete explanation of virtue. Virtue-consequentialisms of the sort I’m interested in must therefore avoid making the right partially explanatory of virtue. Nonetheless, it isn’t obvious that purely goodness-based explanations of virtue cannot capture a strong logical connection between virtue and rightness. This would happen if it turned out that all or most plausible candidates for consequentialist virtues produce good via right actions. In that case virtue-consequentialists could accept the positive motivations for VA in the first two chapters, and welcome at least some of the benefits of VA, while denying that rightness explains virtue, as VA claims. An adequate defense of VA therefore runs though virtue-consequentialism.

I’ll begin by trying to locate the most plausible statement of virtue-consequentialism, construed as a theory about what makes inner states virtuous. It turns out to be surprisingly difficult to locate a statement of virtue-consequentialism that is both true to the defining ideals of consequentialism and respects the distinction between virtue and vice. This gives us significant reason to reject virtue-consequentialism. I then show that no plausible statement of virtue-consequentialism is capable of capturing a strong connection between virtue and rightness. This constitutes another significant mark against virtue-consequentialism, and combines with the earlier difficulties to warrant us altogether to reject the thought that virtue depends solely on a productive relation to the good.

3.2. Stating virtue-consequentialism

Virtue-consequentialisms claim that the virtue of inner states depends on the consequences of those states. Any statement of virtue consequentialism will have to answer three questions. Two of these concern the inner states that function as the theory’s object of assessment. First, exactly what sort of states does the theory take as its object: character traits, occurrent motives, or something else? The second is whether the test for virtue supplied by the theory applies to types or tokens of the state being assessed. Third, virtue-consequentialisms will need to advance an actual consequence-based criterion of virtue. I’ll address these in order.

3.2.1. Traits or motives?

As I understand it, virtue-consequentialism is a theory about what makes inner states virtuous and vicious, but the sort of inner states it takes as its evaluand is open. Perhaps the two most prominent candidates are persistent traits of character and occurrent motives. Which of these we
prefer will depend on our commitments in other areas. The choice may, first, be influenced by our views about the empirical reality of character traits. Recently data from situationist social psychology has been used to call into question the existence of traditional character traits, construed as persistent states that explain and license predictions about behavior across a range of circumstances. Someone convinced by the situationist critique, or even on the fence about it, may prefer to state virtue-consequentialism in terms of occurrent motives, since situationist critiques do not challenge the reality of virtuous occurrent motives.\(^1\)

Second, the choice between traits and occurrent motives might be thought to depend on whether virtuous traits ought to be analyzed in terms of virtuous occurrent motives, or vice-versa. If we think that the virtue of traits is prior to the virtue of occurrent motives, it will be natural to apply the virtue-consequentialist criterion directly to traits, and then to define virtuous motives as whatever motives proceed from, or are characteristic of, the traits the theory deems virtuous. If on the other hand we take virtuous motives to be prior to virtuous traits, we should apply the test for virtue directly to motives, and then define virtuous traits (at least partly) as those that tend to issue in the motives the theory deems virtuous.

To add a further wrinkle, some consequentialists are likely to reject outright the thought that we must choose between traits or motives when deciding where to apply the test for virtue, and to hold instead that virtue-consequentialism ought to assess both of them directly.\(^2\) Just as consequentialists can evaluate rules and acts directly in terms of their consequences, without needing to explain the goodness of one in terms of the goodness of the other, so too, similarly, can they assess the virtue-status of traits and motives.

Rather than attempt to settle these issues, I will simply remain neutral on whether virtue-consequentialism is best stated as a theory about traits or occurrent motives, or both. It seems to me that the difficulties I raise for virtue-consequentialism apply with equal force whether the view takes traits or occurrent motives as its evaluand.

---

1 Hurka 2006
3.2.2. *Types or tokens?*

The second question is whether virtue-consequentialism should assess types or tokens of states. *Token views* assess traits or motives as they are realized in particular people, saying that a particular manifestation of a trait or motive is virtuous just in case it has good (enough) consequences in this instance. *Type views*, on the other hand, apply the virtue-consequentialist criterion to types of states across a population, saying that a type of state is virtuous just in case the consequences of its tokens, taken together, are good (enough). On type views the virtue of particular token states depends on the types of which they are instances. If a token state is of a type deemed virtuous by the relevant criterion, it is virtuous; if not, not.

Importantly, token and type views can yield different conclusions about the virtue- and vice-status of token states. This would happen when a type of state that usually issues in good or bad consequences does not do so in a particular instance. When a type of state that generally has good consequences is tokened in a state that has bad consequences, type views will deem the aberrant state virtuous, since it is a token of a virtuous type; while token views will not, since the instance itself does not have good consequences. *Mutatis mutandis* for vice and bad consequences.

There are abstract considerations that weigh on the choice between type and token views. Julia Driver, for example, endorses type views on the grounds that they minimize the role that moral luck plays in determining which states are virtues. Driver wants to preclude token states that fortuitously produce good consequences from counting as virtuous. On type views these states count as virtuous only if this type of state generally has good consequences; this prevents individual token states from being accidentally virtuous.

However, it isn’t clear whether this solves the problem or just pushes it back a step. While Driver’s strategy may insulate token states from being virtuous by mere luck, it doesn’t seem to insulate type states from being virtuous by luck. Consider a world where *every* token state of some type produces good consequences by bizarre and fortuitous happenstance.³ What is to keep such a type, and thus on type views its tokens, from being virtuous?

Perhaps Driver can reply that such widespread good ‘luck’ would not really be luck at all, but can suffice for a non-lucky connection.\textsuperscript{4} I don’t know whether this can work. There are, at any rate, other considerations that pull against type views.

Driver calls attention to cases where type views yield desirable verdicts about the virtue-status of token states, but in a range of cases this cuts the other way.\textsuperscript{5} Suppose that, as seems likely, the desire to do what we believe to be right (‘conscientiousness’) has good consequences generally, enough to qualify this type of state as virtuous. But now consider Himmler’s desire to do what he thought right; it had extremely evil consequences. Nonetheless, on type views Himmler’s conscientiousness is virtuous, since it is a token of a virtuous type. Here and in other cases type views deem token states with bad consequences virtuous. This is unfortunate for a consequentialist theory of virtue.

The objection here resembles a common worry with rule-consequentialism. Deeming non-optimific acts right just because they issue from optimific rules runs roughshod over the foundational spirit of consequentialism, amounting to rule worship. Analogously, deeming token states with bad consequences virtuous just because they are instances of virtuous types looks like ‘type worship’ and is at odds with the underlying spirit of consequentialism.

A second but related concern involves the difficulty of spelling out the conditions under which a type of state should be regarded as having good consequences, and so being virtuous. One natural thought is that it is the net value of the type’s consequences that matters, such that a type of state has good consequences just in case the net value of the consequences of its tokens is positive. Call this the net value view.

The net value view yields counterintuitive results. Callousness toward the suffering of others almost always has bad consequences. But not unfailingly. Sometimes it has very good consequences, such as when it permits a general to send his troops into a battle that he knows many or most of them will not survive, but which is necessary for the success of some very good

\textsuperscript{4} Since Driver thinks we should evaluate states by considering their consequences at both the world where they are instantiated and other nearby worlds, the state may need to fortuitously produce good consequences at a number of worlds for the objection to stick. But this doesn’t seem impossible.

\textsuperscript{5} Driver 2001: 74f.
end. Imagine now a world where the good consequences of callousness in these exceptional cases outweigh the bad consequences of callousness in all other cases, such that the net value of the consequences of callousness is positive. (A world where, say, several instances of the sort of military callousness illustrated above are needed to prevent domination by an extremely wicked and malevolent force.) On net value views, callousness can then be a virtuous type of state. It would follow, on type views, that callousness is virtuous whenever it’s instantiated (in the relevant world or population). But this is odd. Callousness almost always has bad consequences. Imagine someone (not a military general) who is interested in becoming virtuous. Surely consequentialists should not counsel this person to become callous, or be pleased by his budding apathy!

In light of this it might be thought that a type of state should count as having good consequences just in case the majority of its tokens have good consequences, irrespective of the net value of those consequences. Call this the majority-token view. On this view someone will count as having a virtuous token state just in case most tokens of this type have good consequences.

There are two problems with the majority-token view. The first is that it is still committed to counterintuitive conclusions about token states. Suppose that courage (understood as the ability to stand firm in the face of fear) has good consequences in most cases, but that in some cases it has overwhelmingly bad consequences, such as when it allows military leaders to engage in brutal ethnic cleansing on a widespread scale. On the view in question the courage of these military leaders is virtuous, since courage has good consequences in most other cases. Nonetheless, the courage of these people has horrendous consequences; their courage does not seem good by consequentialist lights.

Second, the majority-token view implies that a type of state that on-balance makes the world worse can count as a virtuous type of state. Imagine that conscientiousness has slightly good consequences in most cases, but extremely bad consequences in a few cases—bad enough to outweigh the good consequences of most of its instances. It follows, on the majority token view, that conscientiousness is a virtuous type of state, since most of its tokens have good consequences. But this again seems unwelcome. A type of state that makes the world on-balance worse should not count as a virtuous type of state by consequentialist lights.
Here we see that type views are caught in something of a dilemma. On the one hand, it seems right by consequentialist lights to restrict virtuous types of states to those that make the world on-balance better. On the other hand, it also seems right to insist that token states with very bad consequences are not virtuous. What the arguments here bring out is that neither net value views nor majority token views can deliver both of these things. Net value views ensure that types of states cannot be virtues without making the world better, but cannot ensure that token states with bad consequences are not virtues. Majority token views at least ensure that most token states produce good consequences, but cannot ensure that consequentialist virtues make the world on-balance better.

Defenders of type views can attempt to respond to these worries by complicating the conditions under which types have good consequences. Perhaps type views should claim that what matters when assessing types of states is not the net value of the consequences of their tokens, but the average value of their tokens.

Note, however, that this will not necessarily preclude counterintuitive results. Suppose again that most instances of some type of state have bad consequences, but that several outlier instances have very good consequences. If the consequences of the outlier instances are good enough, and insofar as the sort of ‘average’ is mean average, the average of the relevant states may be quite high, even if a majority (or indeed nearly all) of their instances have bad consequences.

Type theorists might at this point enlist a different sense of ‘average,’ or look for other ways of stating the view that avoids these results. They might, for example, say that a type counts as having good consequences when both the mean average of the value of its consequences is positive and a majority of the tokens have good consequences. I do not know whether any such statement of type views can, at this point, avoid looking ad hoc. The point is not so much that type statements of virtue-consequentialism are hopeless. It’s rather that they face non-trivial challenges of formulation that they may not be able to meet without giving up a measure of intuitive plausibility.

What about token views? Because token views judge the virtue-status of token states directly on the basis of their consequences, they do not imply counterintuitive conclusions about token states
(from the perspective of consequentialism). They do, however, run into similar difficulties when giving an account of virtuous types.

Even if we have a straightforward way of determining the virtue-status of token states, we may still want to know whether courage or conscientiousness or modesty, construed as general types of states, are virtuous. If token views say that some type of state is virtuous just in case the majority of its tokens are virtuous, we have the earlier worry with type views. It is compatible with conscientiousness having good consequences in a majority of cases that it have on-balance bad, and indeed very bad, consequences overall, in which case the theory will deem a type of state with overall bad consequences virtuous. If on the other hand token views say that some type of state is virtuous just in case the net value of the consequences of its tokens is good, we have the other earlier problem with type views. Callousness might have bad consequences in nearly all cases, but very good consequences in several outlier instances, enough to make the net value of its consequences positive. The view in question would then deem callousness a virtue, despite the fact that it has bad consequences in the vast majority of cases. Token views, then, seem caught in roughly the same dilemma as type views.

Despite the fact that type and token views are subject to the difficulties just discussed, the problems seem to me more worrying for type views. It does not seem as important for consequentialists to be able to give an account of virtuous types as it does for them to give an account of virtuous tokens. We mostly want to know whether the states we actually encounter in our lives, as instantiated by real people—Gandhi’s courage, grandma’s loyalty, etc.—are virtuous.

This thought finds support when we consider whether it would be better, from a consequentialist perspective, for everyone to have the inner states deemed virtuous by a token theory, or for everyone to have the inner states deemed virtuous by a type theory. It seems the former. If everyone has the inner states deemed virtuous by a token theory, everyone will have states that it is good, from a consequentialist perspective, for them to have. If on the other hand everyone has the states deemed virtuous by a type theory, there is no guarantee that virtue will have good consequences in all instances, on either net value or majority token views. But surely from a consequentialist perspective it would be better if everyone had states that directly produced good consequences, rather than states that belong to a virtuous type! This suggests that
consequentialists should focus primarily on evaluating the virtue-status of token states directly, on the basis of their consequences in each instance.

So long as we have a grasp on the virtue-status of token states, determining the virtue-status of types seems less pressing. If it turns out that there is no univocal answer to whether courage considered as a general type is virtuous—if courage has good consequences and so is virtuous in some people, but not in others—this doesn’t seem particularly embarrassing for consequentialists. Just as act-consequentialists will insist that we should focus on the consequences of particular acts, not on whether they are of a type to be ruled out by optimific rules (or whatever), so can virtue-consequentialists insist that it is the consequences of token states that primarily matters.

3.2.3. What is it to have good (enough) consequences?
Virtue-consequentialism claims that a state is virtuous just in case it has good (enough) consequences. The relevant consequences can include the intrinsic value or disvalue of the inner states themselves, but will also include whatever intrinsically good or evil consequences they produce. The most prominent recent statement of virtue-consequentialism, due to Driver, claims that the virtue of inner states depends on their bringing about a net balance of intrinsic good over intrinsic evil.

[T]he account that I want to propose is an objective consequentialist account of the virtues, which would define moral virtues as character traits that systematically produce more actual good than not.⁶

Driver takes traits of character as her evaluand, but her view applies equally well to occurrent motives.

There is an interpretive question with Driver’s statement of virtue-consequentialism, which puts the test for virtue in terms of ‘producing more actual good than not.’ On a strict reading Driver is saying that what is relevant for virtue is just the production of intrinsic goods, such that a state is virtuous iff it produces more intrinsic good than intrinsic evil. This leaves Driver’s statement of the view open to counterexample, since focusing just on the production of intrinsic goods

⁶ Driver 2001: 68.
neglects two ideas at the heart of consequentialism: the positive importance of preventing evil, and the negative importance of preventing good.

Ben Bradley assumes the strict interpretation of Driver’s view, and presses this worry with the case of Downer.

Consider a person named ‘Downer’ who lives in a world full of happy people. Downer engages in various activities designed to lessen their enjoyment; he replaces the pleasures they would have otherwise had with less pleasant pleasures. For example, he invents new, less tasty foods and hoodwinks those people into thinking those foods are healthier than the tasty foods they like better. He never causes them any pain—he does not have the stomach actually to hurt anyone—but he does cause them to be less happy than they would have been otherwise. There are no intrinsically bad consequences to appeal to here. Downer brings a lot of intrinsic good into the world, consisting of all the moderately pleasant experiences people have as a result of his actions (Bradley 2005: 284).

If all that matters to virtue-status is the production of intrinsic goods, Downer’s trait is a virtue, since it brings about a net balance of intrinsic good. This is, Bradley claims, the wrong conclusion for a consequentialist theory of virtue to return. Downer sees to it that people are less well-off than they would’ve otherwise been. Since consequentialism is, or should be, concerned with good prevented as well as good realized, a consequentialist theory of virtue should not imply that Downer is virtuous.

Worse still, it seems to me, is the fact that focusing just on the production of intrinsic goods neglects the importance of preventing intrinsic evils. Consider someone who produces no intrinsic goods, but prevents significant amounts of suffering, leaving the world much better than it would otherwise be. You are a military medic equipped with painkiller, so motivated as to spend each day administering this painkiller to wounded and dying soldiers in the field. We can imagine that the painkiller alleviates pain without producing any euphoria or pleasure, in which

---

Bradley also seems to recognize the importance of preventing evil to a consequentialist theory of virtue, even if the Downer example doesn’t bring it out. He says that reflection on Downer “sheds light on an important consequentialist idea: preventing goodness is just as bad as producing evil, and preventing evil is just as good as producing goodness” (2005: 285).
case it prevents evil while producing no intrinsic goods. Since you have produced no intrinsic goods, you would not count as virtuous on the strict interpretation of Driver’s view. But given that you have prevented much suffering, your motives should be virtuous by consequentialist lights.  

Note that a similar example can be used to show that net production of intrinsic evil does not qualify a trait or motive as vicious, or disqualify it as virtuous. Perhaps the painkiller in question has negative side-effects, causing some slight suffering while nonetheless leaving the patient better off than she would otherwise be. Or imagine a dedicated trolley-diverter, someone who travels around from line to line diverting trolleys headed toward five people on to tracks containing one. This person causes intrinsic evil, by causing the death of the one in each case. Nonetheless, she prevents much more evil, by sparing the five. Such a motive or trait, one that prevents significant suffering, should count as a consequentialist virtue, even if it produces a net balance of intrinsic evils.

Partly because of these considerations, it seems to me uncharitable to claim that Driver’s intended view is that virtue depends solely on the production of intrinsic goods. The main textual evidence Bradley supplies for the strict interpretation is the following sentence of Driver’s: “On this view, with respect to character traits, one is comparing the good produced with the bad produced and judging the good to be greater.”

Driver does seem to be giving us a test for when a state’s consequences produce more intrinsic good than intrinsic evil at some world, for the end of assessing its status as virtuous. But I do not see her claiming that this is the only thing that matters for virtue-status. Indeed Driver repeatedly speaks of the virtues in terms that implicitly acknowledge the importance of decreasing or preventing evil.

---

8 Norcross 1997 used similar considerations in the context of examining and rejecting typical consequentialist approaches to good and bad actions. Many of the same considerations apply to consequentialist evaluations of motives and traits.

9 I owe the trolley-diverter example to Brian Berkey, who provided valuable comments on the core ideas of this chapter at the 2009 Pacific APA.

On the account I offer, virtues function in social contexts to contribute to human (or social) flourishing and happiness, often by alleviating interaction problems among people.\textsuperscript{11}

For example, trustworthiness and honesty make social interaction feasible, since without these traits it would be very difficult for people to coordinate their activities. They make interaction much more efficient. They cut security costs.\textsuperscript{12}

Driver also explicitly compares virtues to seat belts, claiming that “the virtue of the seat belt is that it saves lives,” as well as to sprinkler systems in buildings, whose function is to prevent damage from fire.\textsuperscript{13} These analogies support the idea that, for Driver, an important function of virtue is to prevent evil, not merely to produce good.

\textit{Pace} Bradley, then, it seems to me that when Driver says that virtues must ‘produce more actual good than not’ this is simply another (perhaps less than maximally precise) way of saying that states must have good consequences in order to be virtues. In that case Driver’s view need not be subject to the charge that it ignores the negative import of preventing good and the positive import of preventing evil, since ‘having good consequences’ can be taken to be a function of both of these things.

Of course, this still leaves Driver and virtue-consequentialists generally with the task of saying precisely what is required for a state to have good enough consequences to count as a virtue. Part of what the preceding examples bring out is that, when we say that some state has good or bad consequences, we are making an essentially comparative claim. We are saying that because of the state in question the world is better or worse. Better or worse than what? To say that state $s$ makes our world better is presumably to say that the actual world is, by virtue of containing $s$, more valuable (or less disvaluable) than one or more worlds that lack $s$. But \textit{which} worlds that lack $s$? There are many ways for worlds to lack $s$.

Consider a token state of moderate generosity, instantiated at the actual world by Sam. To say that Sam’s generosity has good consequences is to say that the world would be worse were Sam

\textsuperscript{11} Driver 2001: 74, my emphasis.
\textsuperscript{12} Ibid, my emphasis.
\textsuperscript{13} Ibid., pp. 74-75.
not in this state. One way for this to happen is if Sam were not around to be generous: if Sam were not in existence to instantiate the trait. This, however, complicates things toward futility, since now we somehow need to distill the difference Sam’s generosity makes between worlds from the difference the rest of his existence makes between worlds.

Even if we keep Sam on the scene there are many ways for him to fail to be generous. Sam could fail to be moderately generous by having states with better consequences—by being a complete saint. Or he could fail to be generous by having states with worse consequences—by being avaricious or apathetic or positively spiteful. Whether or not Sam’s generosity makes the world better depends on which of these alternatives we measure it against. If we compare our world, where Sam is mildly generous, to the world at which Sam is extremely generous (holding everything else as equal as possible), Sam’s mild generosity does not make our world better. If on the other hand we compare Sam’s mild generosity to the world where Sam is malicious, it does make our world better. We thus need to know what the appropriate comparison is before we can determine whether states make the world better or worse, and so before we can deem any state virtuous or vicious.

One straightforward strategy is to identify the baseline as what the world would be like if it lacked the state in question. When we want to decide whether Sam’s mild generosity has good consequences, we compare the actual world, where Sam is generous, to the closest world where Sam is not. If the closest world where Sam lacks generosity is worse (or worse enough) than the actual world, Sam’s generosity is virtuous at our world; if the closest world where Sam lacks generosity is better (or better enough), Sam’s generosity is not virtuous (and may be vicious) at our world. The relative baseline, when measuring the consequences of states at a world, is just the nearest world at which those states aren’t instantiated.

Bradley rejects this strategy on the grounds that there is no univocal answer to what worlds are closest. But even putting aside this difficulty, the view returns unacceptable conclusions. We can see this by considering states that produce significant evil but keep the world from being worse, or states that produce significant good but keep the world from being better.

---

Imagine that Millie is spiteful in the actual world, causing much suffering. It might nonetheless be true that if Mille were not spiteful she would be even worse (say, unrestrainedly malevolent), producing even more suffering. Thus, Millie’s spitefulness has good consequences by the relevant test, and may count as a virtue. But Millie’s spitefulness causes much suffering; it should not count as a virtue by consequentialist lights.

Or imagine that Jon is mildly compassionate at our world, bringing about significant good. Suppose, however, that Jon would be even better were he not mildly compassionate (say, totally selflessly loving), bringing about even more good. Then Jon’s mild compassion has bad consequences by the relevant test, and will not be a virtue. Again, this is the wrong result for a consequentialist theory to return. It looks like the closest world/s test does not provide an attractive statement of the view.

Perhaps we can make progress here by focusing on more familiar versions of act-consequentialism. When act-consequentialists want to know whether some act is right, they compare it to other acts that could have been done instead. Similarly, when considering whether a token state makes our world better, it is natural to ask whether this state has better consequences than other states this person could have had (not necessarily would have had) instead.¹⁵

What is the sense of this ‘could have had’? Presumably it is something stronger than epistemic possibility. When assessing Sam’s generosity we aren’t interested in measuring it against other states that we can merely imagine Sam having. Presumably we want to know whether the states Sam has are better or worse than other states that are (or were, at some point) real possibilities for Sam—states Sam may well have realized. One important question here is whether the states in the contrast class, those we ‘could have had,’ are states we could have had given our actual mental and physical constitutions, and the conditions into which we are born; or whether these things should not place limits on the contrast class. I’m inclined toward limiting the contrast class to those states we could have had given our actual constitution and circumstances. These states will form the contrast class against which we measure Sam’s actual generosity.

Suppose, then, we know which states are in the contrast class, and how the state we are assessing measures up to them. We have an ordinal ranking of all the states Sam could have had, including the one he does have, generosity. We can then say how much better or worse Sam’s generosity is than the other states he could have had instead.

Unfortunately, this does not yet tell us whether Sam’s generosity, or any other token state we wish to assess, is virtuous, since it isn’t obvious how to move from an ordinal ranking of this sort to conclusions about the virtue- or vice-status of states on the list. Perhaps the most straightforward view is a maximizing one. It identifies virtue with the best state on the list, and vice with the worst. This would amount to saying that a given state is virtuous just in case it’s the very best state the person in question could have, and that a state is vicious just in case it’s the worst. But a maximizing interpretation of virtue-consequentialism is implausible, for at least two reasons.

First, it’s overly demanding to insist that virtues must be the best states we could have had, and vices the worst.¹⁶ Imagine that Sam, because of his generosity, causes significant good in the actual world. Suppose nonetheless that Sam could have possessed a state with even better consequences (say, unrestrained universal benevolence). Then, on a maximizing view, Sam’s generosity will not count as virtuous. But this isn’t right. Even if unrestrained universal benevolence is more virtuous, being generous is still virtuous.¹⁷ Similarly, someone who is in fact spiteful, but who could have been even worse, seems no less vicious for that.

Second, the maximizing view returns the wrong conclusions in cases where the facts put constraints on what states we could have had. Since on the view under consideration virtuous states are the best states that someone could have had, some people may be so unfortunately constituted and/or situated that all of the states that they could have—and so the best state they could have—produce significant evil.

¹⁶ Driver 2001: 73.
¹⁷ Similar points apply, mutatis mutandis, for restricting vices to the worst states.
Imagine that Theresa is spiteful, and that her spitefulness is the cause of significant pain and suffering in the actual world. But now suppose that, given vagaries of constitution and circumstance, this is, tragically, the best Theresa could hope to be (in this respect). If so, Theresa’s spitefulness is, on maximizing views, virtuous. But the fact that Theresa could not be better in this regard should not imply that her spitefulness, which after all brings about significant evil, is virtuous (though it may influence our willingness to blame her for it).

Conversely, agents may be so fortunately constituted and/or situated that all of the states that they could have—and so the worst state they could have had—produce significant good. Imagine that Alice is mildly compassionate, and that her compassion produces happiness in our world. Unlike Theresa, Alice is lucky. Given her fortunate constitution and upbringing, mildly compassionate is the worst she could have been. Here maximizing views imply that Alice’s compassion is a vice, since it is the worst state she could have. Again, however, the mere fact that Alice could have had better states should not imply that Alice’s compassion, which brings about significant happiness, is vicious.

This last objection may, I admit, turn on controversial views about what states someone ‘could have had.’ Someone who takes a liberal approach to ‘could have had’ can perhaps reject the thought that the range of alternate states is sometimes restricted in the way the objection assumes. But even if this is so, the objection points to a deeper problem. Sometimes, at least, what other states we could have had simply does not seem relevant for whether the states we do have count as virtues or vices. I return to this below.

What other options are there for stating the view? Bradley, for his part, claims that there is no non-relative answer to whether a motive or trait is virtuous.

No character trait has the property of being a virtue full stop, because there is no such property. Instead of attributing such a property to a trait, we should say that it is a virtue to have one character trait C1 rather than some other trait C2. When we call a character trait a virtue, context must provide another character trait to fill in the second place of the relation (2005: 286).

Bradley thus claims that states are only virtuous or vicious relative to other states in the contrast class. Take some inner state $s$ whose virtue- or vice-status we wish to assess. Relative to those
states in the contrast class whose consequences are worse than \( s \), \( s \) is a virtue. Relative to those states in the contrast class whose consequences are better than \( s \), \( s \) is not a virtue and indeed is a vice.\(^{18}\)

So if we want to know whether Reba’s actual tendency to waver but stand her ground in the face of danger is virtuous, we compare this to other states Reba could have had, say, the tendency to run away as fast as she can in the face of danger; calmly to stand firm in the face of danger; and so on. If Reba’s tendency to waver but stand her ground has better consequences than a tendency to run away, the former is a virtue relative to the latter. If on the other hand the tendency to waver but stand her ground has worse consequences than a tendency calmly to stand firm, the former is a vice as compared to the latter. Reba’s tendency to waver but stand her ground is a virtue relative to some ways she could have been, and a vice relative to some other ways she could have been. While it may look like we predicate virtue and vice univocally, the appearance is misleading. Properly understood, virtue-ascriptions are always disguised comparisons between states. There is simply no answer as to whether a state is virtuous or vicious simpliciter.

Bradley’s contrastivism is subject to several difficulties. First, the view is too narrow, precluding states that seem to be virtuous from counting as such. Consider again Sam, whose generosity brings about good. Suppose again that Sam could have possessed an even better motive, (say) the motive of selfless love, with better consequences. According to Bradley, Sam’s generous motives are not virtuous, and indeed are vicious, relative to the motive of selfless love. But this isn’t right. While selfless love may be more virtuous than generosity, this should not imply that it is vicious to be generous rather than selflessly loving.

Second, Bradley’s contrastivism is too wide, counting states that seem to be vices as virtues. Consider again Millie, whose spitefulness causes suffering in the actual world. Imagine again that Millie could have possessed an even worse motive than spitefulness, say, the motive of unbridled malevolence. Then it will be true that spitefulness is a virtue relative to unbridled

\(^{18}\) “The trait of being fairly honest has good consequences, and so counts as a virtue, when compared to the trait of being mostly dishonest … it has bad consequences, and counts as a vice, when compared to the trait of being mostly honest” (Bradley 2005: 286).
malevolence. But this isn’t right. While spitefulness may be less vicious than unbridled malevolence, it is not virtuous for Millie to be spiteful rather than malevolent.

These difficulties stem from Bradley’s claim that virtue-judgments are always and only hidden comparisons. No doubt ‘virtue’ and ‘vice’ can be employed comparatively; it is possible to be more or less virtuous, and more or less vicious. But virtue and vice concepts also seem to function in a more absolute sense, as threshold concepts. To see how virtue and vice can be both threshold and comparative notions, consider the idea of supererogation.

There is a certain threshold of goodness (or whatever) that acts must meet to be supererogatory. If they do not meet this threshold, they will be obligatory, not supererogatory, and no comparative supererogatory judgments will apply to them. Once they cross the relevant threshold, however, we can compare them, asking how supererogatory each act is. (This simply amounts to asking how far beyond the call of duty each one goes.) The same goes for virtue and vice. The intuitive view is that a state must measure up to a relevant bar to be virtuous. The states that do measure up are virtuous, full stop; and the states that do not measure up aren’t virtuous, full stop. But at the same time we can assess states that do meet the bar as more or less virtuous than each other. (Mutatis mutandis for vice.) In this sense virtue and vice are both threshold and comparative notions.

The question for virtue contrastivists is whether denying that virtue and vice are threshold notions, in the sense just sketched, is an acceptable cost to bear. We do, I think, have a strong sense that virtue and vice are threshold notions. This explains why it is difficult to accept that it is virtuous for Billy to be spiteful rather than unrestrainedly malevolent. No matter what other motives Billy could’ve had, Billy seems vicious for being spiteful. Similarly, no matter what other motives are available to us, the motives of generosity or courage or justice—the motives of Saint Francis or Rosa Parks or Helen Keller—seem virtuous (full stop). A theory of virtue should, in other words, capture more than the thought that Mother Theresa was better than she

---

19 I think we could also simply consider the concepts of ‘good’ and ‘bad.’ Since, however, some people appear to think that good and bad are only comparative notions, appealing to supererogation may be less controversial; I don’t know anyone who claims that the concept of supererogation functions only comparatively.
could have been, or Stalin worse than he could have been. It should let us say that she was virtuous, and he vicious, *simpliciter*.

Bradley is likely to reply that this amounts to begging the question against the view. We would then have to quarrel over whether the counterexamples raised here are ultimately reasonable bullets to bite. There is, however, a related but distinct objection to Bradley’s contrastivism, one that goes out of its way to engage it on its own turf.

Bradley explicitly endorses the view that there are no absolute or non-relative virtues or vices. But even if we are willing to accept that states are never virtuous or vicious *simpliciter*, we might still, on Bradley’s view, hope to be able to say that states can be more or less virtuous, or more or less vicious, than others. At first glance Bradley’s view seems well-suited to capture these judgments. After all, Bradley just seems to be saying that, so far as virtue-judgments, all we have are claims about which states are better and which worse, and so more or less virtuous or vicious, than others.

Given the details of Bradley’s view, however, it’s not at all clear how to move from claims about the comparative value of states (‘better than’; ‘worse than’) to conclusions about their relative virtuousness and viciousness (‘more/less virtuous’; ‘more/less vicious’).

Imagine again the trolley-diverter. Suppose motive *m* is responsible for the trolley diverter saving many lives, and that *m* is a better motive than ten of the motives in the relevant contrast class, but a worse motive than one, *n*. On Bradley’s view it seems correct to say that, in having *m*, the trolley-diverter is more virtuous than she could have been, since there are ten other worse motives, relative to which *m* is virtuous, that she could have had instead. By the same token, we should also say that, in having *m*, the trolley diverter is more vicious than she could have been. Relative to *n*, after all, *m* is vicious. In that case Bradley’s contrastivism deems the trolley-diverter’s motives both more virtuous and more vicious than they could have been.

This is intuitively unwelcome. While the trolley-diverter’s motives may be less virtuous than *n*, they do not seem more vicious than *n*. To take a more drastic case, suppose that Jackie Robinson or Martin Luther King, or your favorite paragon of virtue, in fact had motives much better than many other motives they could have had instead, but not the very best motives they could have had. If so, Bradley’s view entails that the admirable motives had by Robinson and
King are *more vicious* than other motives, the best, they could have had instead. But while it may be true that the motives of Robinson and King were less virtuous than they could have been, they are not more vicious. Likewise, someone who is spitefully motivated, but not as horrendously motivated as she could be, is less vicious than she could have been, not more virtuous.

In addition to precluding non-relative attributions of virtue and vice, then, Bradley’s view also commits us to saying that people with very good states, but not the very best, are more vicious (not just less virtuous) than they could have been; and that people with very bad states, but not the absolute worst, are more virtuous (not just less vicious) than they could have been. This is a direct result of denying that virtue and vice are threshold concepts, and a second price to pay for the view.

It is, moreover, a fairly steep price. The mere fact that some motive is not the best we could have should not imply that we are more vicious than we could have been. It would be manifestly absurd to insist that Jackie Robinson’s courage made him less vicious, because he could have been even better! There is nothing vicious at all about his courage. Likewise, the mere fact that someone could have been worse should not imply that he is more virtuous for that. Iago’s spite did not make him more virtuous, merely because he could have been worse. There is, again, nothing virtuous at all about his spite.

In other words, it seems plainly false that being less virtuous is the same as, or equivalent to, being more vicious; and no less false that being more virtuous is the same as, or equivalent to, being less vicious. If on Bradley’s view there is no way to distinguish being more virtuous from being less vicious, or being less virtuous from more vicious, the distinction between virtue and vice is in trouble. This is a further significant mark against Bradley’s contrastivism, one that can combine with the earlier difficulties to warrant us to reject it altogether.\(^\text{20}\)

\(^{20}\) Note that appealing to a type statement of Bradley’s contrastivism will not help with these objections. Presumably type theories will, like token theories, begin by delimiting an ordinally ranked contrast class of states, against which particular types of states can be assessed. Here type views encounter the same problems moving from the list to conclusions about the virtue- and vice-status of states on it.
Is it possible to develop the basic idea behind virtue-consequentialism—that the virtue of inner states is determined by the value-difference they make to a world—in a way that can escape these difficulties? Perhaps the most natural way to resist the objections so far would be to try to locate what Bradley calls a ‘benchmark’ state from among the contrast class. States with better consequences than the benchmark state will be virtuous, while states with worse consequences will be vicious. (Or, if one is concerned that this entails that there are no neutral states—that every state must be either virtuous or vicious—one could posit two benchmarks, one for virtue and one for vice, with space in-between.) This would be to recognize the nature of virtue and vice as threshold statuses. In so doing it would allow us to escape the previous worries.

Bradley rejects this strategy on the grounds that any selection of the benchmark will be arbitrary.\textsuperscript{21} But this might be too quick. Indeed it might be thought that there is a principled and relatively straightforward way to determine a benchmark state. Suppose we are assessing Sam’s generosity, where ‘generosity’ denotes a desire whose content is to help others in need. The challenge is to specify a principled baseline against which we can determine whether this desire has good or bad consequences. Bradley assumes that, when assessing this desire, we can measure it against any desire that Sam could have had instead, and indeed that it is just as appropriate to measure it against any one state in the contrast class as any other. But why not think that the most appropriate baseline is simply the absence of the desire in question?

Imagining a world where Sam has no desire to help. It is not that Sam has a positive desire to harm at this world. Rather, he simply has no concern either way about the well-being of his fellows. If this is a state Sam could have had, why should it not be the baseline against which the consequences of his generosity are judged? Call the absence of the desire we are assessing the ‘null state.’ The strategy on offer then says that, in general, the appropriate baseline for evaluating the consequences of any state is the null state, which provides the benchmark for whether states are virtuous are vicious. States with better consequences than the null state are virtuous, while states with worse consequences are vicious. Call this the null state view.\textsuperscript{22}

On its face the null state view is appealing. However, it faces its own set of challenges. Perhaps the biggest worry is whether it can avoid smuggling in substantive assumptions about the virtue

\textsuperscript{21} Bradley 2005: 287.
\textsuperscript{22} I owe the null state strategy to Tom Hurka.
and vice-status of particular states. It should, after all, be an open question whether a null state itself is virtuous or vicious. If a null state has good consequences, it should be virtuous; if it has bad consequences, it should be vicious.

But how are we to measure the consequences of null states, if the null state always functions as the baseline? We cannot measure the consequences of the null state against itself; doing so yields the conclusion that lacking certain states never has good or bad consequences, and so is never virtuous or vicious. This seems false. It is intuitively false insofar as lacking the desire to help someone in dire need (for example) will often strike many of us as vicious, rather than neutral. Moreover, often someone who lacks the desire to help others will, as a result of this lack, fail to help them. In so doing they may often make the world worse.

Of course, whether lacking the desire to help has good or bad consequences depends on what we measure it against. But now we are back to the same difficulty as before: specifying the correct baseline. The null state view was meant to provide a solution to the baseline problem. Its answer was that we should always measure the consequences of the state we are evaluating against the consequences of the null state. What this objection brings out is that we cannot discern the consequences of the null state without specifying a baseline that is not the null state. The null state view thus looks to be incapable of assessing the virtuousness and viciousness of lacking states, which is a significant mark against it.

Here is a second worry with the null state view. Notice, first, that the world at which a person has a null state may be further (indeed much further) away from the world we are evaluating than other worlds where that person has other states instead. This can make the null state view look arbitrary. We can see this by focusing on a situation where a state has better consequences than other states that are instantiated in its place at most other worlds, but worse consequences than the null state.

Suppose that Tina has a mild desire for wealth and material luxury. At the closest world where Tina lacks this desire, she has a desire with worse consequences—say, utter avariciousness. Moreover, at most worlds Tina’s desire for wealth is too strong, leading to worse results than her actual mild desire. The only way Tina could have had a better state is, let’s suppose, to have the null state: simply to lack the desire for wealth. But the world at which Tina has the null state is
(ex hypothesi) very far away from the actual world, and a very unlikely possibility. In this case Tina’s mild desire for wealth keeps the world from being worse, and perhaps significantly worse, than it would be under nearly all of the alternative ways things might have gone instead. This makes it look like a good candidate for a consequentialist virtue. However, since Tina’s mild desire for wealth has worse consequences than the null state, it is not a virtue, and may be a vice, on null state views. This seems ad hoc. Why should we measure Tina’s actual desire against such a far-flung possibility? The null state view struggles to answer this question.

I’ve been arguing that a myriad of obstacles confront attempts to formulate a satisfactory version of virtue-consequentialism, one that captures the underlying spirit of consequentialism while refraining from delivering unacceptable conclusions about the virtue- and vice-status of particular states. Perhaps the counterintuitive implications of virtue-consequentialism are unsurprising, given the well-recognized extent to which traditional act-consequentialism is at odds with our considered moral judgments about the deontic status of acts. Of course, just as act-consequentialists have worked hard to meet these criticisms, virtue-consequentialists can attempt to overcome the difficulties raised in this section. Perhaps more likely, die-hard virtue-consequentialists can simply stress the radical and revisionary nature of consequential theorizing. Because this seems to me the likely response from many consequentialists, I will use the rest of the chapter to point out a different and independent set of problems with virtue-consequentialism, which can combine with the shortcoming exposed in this section to further ground skepticism about the view.

3.3. Virtue-consequentialism, right action

In chapter one I appealed to intuitively plausible claims about the abstract relations between virtue and other moral concepts to motivate VA over the Prichard-Ross approach. In particular I argued that VA does a better job than P-R of capturing relations between virtue and benefit, both individual and social. As I mentioned then, virtue-consequentialisms have an edge over VA when it comes to explaining the social benefits of virtue. This is because virtue-consequentialisms can presumably explain all of the social benefits of virtue, while VA can explain only those social benefits that come about by means of objectively right action.
Whether virtue-consequentialism has an edge over VA when it comes to explaining the individual benefits of virtue is more complicated. I argued in chapter one that the individual benefits of virtue involve the goods of well-motivated objectively right action. For virtue-consequentialism to explain these benefits, it must ensure a strong connection between being virtuous and doing the objectively right thing. So whether virtue-consequentialism provides a better explanation of the individual benefits of virtue thus turns on the set of claims it implies about the relation between virtue and the right. In this section I’ll argue that virtue-consequentialism does not secure a plausible connection between virtue and the right. This in itself will, I think, give many reason to reject it, given the virtue-tradition’s insistence that virtue and right action are closely linked. It will also show that, unlike VA, virtue-consequentialism cannot capture the benefits that accrue to individuals who perform well-motivated objectively right acts beneath the concept of ‘virtue.’

3.3.1. Goodness-based, rightness-based

It might seem easy for virtue-consequentialists to guarantee a strong connection between virtue and right action. Consider G.E. Moore, himself no consequentialist slouch.

The test for the ethical connotation of virtue is the same as that for duty: What should we require to be proved about a particular instance, in order to say that the name was wrongly applied to it? And the test which is thus applied, to both virtues and duties, and considered to be final, is the question: Is it a means to good? If it could be shewn of any particular disposition, commonly considered virtuous, that it was generally harmful, we should at once say: Then it is not really virtuous. Accordingly a virtue may be defined as an habitual disposition to perform certain actions, which generally produce the best possible results. Nor is there any doubt as to the kind of actions which it is ‘virtuous’ habitually to perform. They are, in general, those which are duties ...

This is an interesting passage. Moore first, like Bentham and Hume before him, makes instrumental goodness the test of virtue: “Is it a means to good?” But then Moore explicitly defines virtue as a disposition to perform acts of a particular sort—those “which generally produce the best possible results”—viz., consequentialist right actions. In doing so he seems to

---

23 Moore 1903 (1993), section 103, p. 221.
be explaining virtue in terms of right action. Since right action is for Moore a function of instrumental goodness, these two claims are compatible. Moore seems to be saying that a state counts as a virtue just in case it produces instrumental value by way of consequentialist right action.

Why, however, does Moore restrict the ways in which states can realize good consequences to the production of right action? This is peculiar. For one, as I discuss below, recent authors have shown that a state’s consequences can outstrip the consequences of the acts they explain, for example, by directly producing pleasure in their possessors, or being themselves intrinsically good or evil. But even if Moore missed this, surely he would have seen that states can produce good consequences by producing wrong acts that nonetheless have good consequences; that is, by producing second or third or fourth best acts, by consequentialist lights.

Because I doubt Moore missed this, I think he is best read in this passage as placing an explicit right action constraint on virtue. This, however, threatens to disqualify him as a virtue-consequentialist. If a state must produce (consequentialist) right actions to count as virtuous, we need to ask whether Moore’s theory isn’t at least partly rightness-based, rather than solely goodness based.24 Since I take ‘virtue-consequentialism’ to refer to views on which virtue depends just on goodness produced, with no reference to rightness, this amounts to asking whether a view like Moore’s is a true version of virtue-consequentialism, rather than a mixed theory that makes virtue depend on both the good and the right.

What makes Moore’s view interesting from a consequentialist perspective is that he seems to place a special requirement on the way in which the good consequences of inner states must be realized in order for them to count as virtuous; namely, via consequentialist right action. Of course, this distinguishes Moore’s view from virtue-consequentialism only if it’s possible for inner states to produce enough good consequences to count as virtuous without producing consequentialist right actions. If standard virtue-consequentialisms wind up entailing that virtuous inner states produce consequentialist right action, there would be nothing special, from a virtue-consequentialist perspective, about Moore’s theory.

---

24 It is likely that Mill and Sidgwick as well held a view like Moore’s; see below.
If this turns out true—if virtue-consequentialism entails that virtuous states of necessity produce right acts—it becomes a tricky question whether virtue-consequentialism would still count as a purely goodness-based, rather than partly rightness-based, view—one that turns on the relation between logical entailment and metaphysical explanation. Since it can be denied that $p$’s logically entailing $q$ entails that $q$ metaphysically depends on $p$, it seems possible to hold that the presence of virtue entails the existence of right action while denying that right action explains virtue.\(^{25}\) If so, it would be possible for virtue-consequentialists to claim a logically necessary connection between virtue and right action while also insisting that their theory is purely goodness-based, in which case Moore’s view could count as a pure form of virtue-consequentialism.

If on the other hand virtue-consequentialism entails that it’s possible for inner states to count as virtues without producing consequentialist right action, Moore’s view looks like a substantive alternative to virtue-consequentialism. In fact I think this latter scenario is how things stand. To show this I’ll argue that it is possible for consequentialist virtues to consistently and indeed always issue in wrong acts.

3.3.2. *Rightness and consequence*

On virtue-consequentialism, the fact that an inner state $s$ is virtuous does not entail that $s$ issues in right action. This is so whether the notion of right action in question is subjective or objective. Let’s start with subjective rightness. There is no guarantee that a state that produces very good or even the best consequences will produce acts that are subjectively right, or right from the perspective of the agent, since states can produce subjectively wrong acts with very good consequences.

Sam needs the antidote, and Fran must decide whether to give him the medicine in the green bottle, or the medicine in the red bottle. Imagine that the subjectively right thing for Fran to do is to give Sam the antidote in the red bottle; Fran’s beliefs support the conclusion that the antidote in the red bottle will cure Sam, and Fran believes as much.\(^{26}\) Fran, however, does not

---

\(^{25}\) Cf. VA to virtue-based theories of right action, which try to explain right action in terms of virtuous action (see chapter four). Both claim that wherever there is virtuous action there is right action. The dispute is over whether the rightness or virtue of acts is more basic.

\(^{26}\) We can also imagine that these beliefs of Fran are justified, if we need to.
like Sam, and wants him to suffer. So she gives Sam the antidote in the green bottle, thinking it won’t work. Fortunately for Sam, Fran is wrong about which antidote will cure him; the antidote in the green bottle is the right one, and Sam is cured.

In this case Fran’s desire to harm Sam has very good consequences, since it causes Fran to give Sam the antidote that saves his life. If these consequences are good enough, as they may well be, Fran’s desire will meet the consequentialist bar for virtue. At the same time, Fran’s desire causes her to do what’s subjectively wrong. Thus, consequentialist virtue does not guarantee subjectively right acts. This is at least partly because there is never any guarantee that the agent’s beliefs, those relevant for determining subjective rightness, will be true, or support the act that in fact has good consequences. If we are bad at tracking relevant facts, our beliefs might often instruct us to perform acts with bad consequences. In these cases the inner states that have or would have good consequences are not those that would cause us to do what’s subjectively right.

What about the relation between consequentialist virtue and objective rightness? Here the argument takes a bit more work. For the purpose of the objection we should assume a consequentialist standard of act evaluation. The question is whether virtue-consequentialism can ensure a strong connection between virtue and consequentialist right acts.\(^{27}\)

Consider some inner state \(s\) that brings about lots of good in the actual world, and so is a promising candidate for a consequentialist virtue. Suppose first that the good consequences \(s\) brings about are mediated entirely via the consequences of the acts \(s\) explains, considered in abstraction from \(s\). Also suppose that traditional (maximizing) act-consequentialism is correct: acts are right just in case they bring about the best consequences of any alternative in the situation.

It’s compatible with \(s\) producing large quantities of good, via the consequences of the acts it explains, that \(s\) produce acts which fail to be right. This would happen if \(s\) produced acts that brought about significant amounts of good without bringing about the most good. If the non-optimal acts \(s\) produces are valuable enough, \(s\) can produce enough good consequences to meet a

\(^{27}\) It is easier to show that virtue-consequentialism does not ensure a strong connection between virtue and rightness if act-consequentialism is not correct.
reasonable bar for virtue. In that case virtue fails to guarantee right action, and indeed can always lead its possessor to do the wrong thing.

Someone might object that this is so only if we assume the notoriously over-demanding criterion of act-evaluation endorsed by maximizing act-consequentialism. It seems true that this line of thought gets trickier to run if we complicate things with different consequentialist standards of act and virtue evaluation. If we raise the bar for virtue and lower the bar for acceptable action, for example, it doesn’t seem nearly so straightforward that a state could have good enough consequences to be virtuous without producing right action. There is, however, an independent set of considerations that show that consequentialist virtues can always issue in wrong action.

Several have pointed out that the value-difference made to a world by the presence of certain character traits or motives isn’t exhausted by the value-difference made to a world by the presence of the actions those traits or motives explain. Robert Adams has something like this in mind when he says: “Even if there is no difference in external circumstances, the motivational pattern that leads to more useful actions is not necessarily the more useful of two motivational patterns, on the whole.”28 Derek Parfit appears to endorse something similar, as do Phillip Pettit and Michael Smith.29 I’ll call the claim that the consequences of a trait or motive aren’t exhausted by the consequences of the actions they explain the hypothesis of Non-Exhaustion.

*Non-exhaustion:* The value-difference made to a world by an inner state is not exhausted by the value of the consequences of the acts it explains.

If non-exhaustion is true, it will always be possible, in principle, for states that bring about wrong acts to be virtuous. Even if the consequences of the acts produced by some state aren’t at all impressive, there will, if non-exhaustion is true, be other ways for the state to produce enough good to count as virtuous. The status of the hypothesis of non-exhaustion is thus highly relevant for determining the relation between right action and consequentialist virtue.

---

3.3.3. **Non-exhaustion**

There are at least three ways of arguing for non-exhaustion, each of which seem promising. The first calls attention to the virtues of restraint: temperance, patience, humility, and the like. One main function of the virtues of restraint is to prevent evil, by keeping us from acting in bad or harmful ways.\textsuperscript{30} Suppose that Tim is temperate. It is likely that Tim’s temperance will make the world better by keeping him from being overly tempted by immoderate pleasures. In order to support non-exhaustion, it must be true that Tim’s temperance has good consequences apart from any of the actions it explains. But this seems plausible. If Tim’s temperance keeps him from doing things he shouldn’t, it seems good because of the actions it keeps him from doing, not because of the actions it produces.

Someone might yet worry that the virtues of restraint keep us from doing the things we shouldn’t by allowing us to act in ways we wouldn’t otherwise have acted. Temperance is good when it allows us to choose a glass of water before bed, rather than succumb to a milkshake. Patience is good when it allows us to turn slowly and walk away, rather than say something we’ll regret. Here the virtues of restraint keep the world from being worse by causing us to perform better acts than we would have performed. In that case they would not support non-exhaustion.

However, there also appear to be times when the virtues of restraint keep the world from being worse simply by making it the case that we refrain from acting. For this to work it must be true that the omissions explained by the state in question are not actions in the relevant sense.\textsuperscript{31} But I see no reason to think that every time temperance or patience keeps us from doing things we shouldn’t it does so by enabling an intentional omission. It is not that Tim must consciously focus on refraining from immoderate pleasure every time his temperance keeps him from pursuing it. Rather, Tim’s temperance might keep him from succumbing to milkshakes simply by making it true that they aren’t a real temptation for him in the first place.

\textsuperscript{30} Kant seemed to think that it is of the very essence of virtue to restrain us from being led astray by our inclinations (MM IX, 6:394-5). See also Roberts 1984.

\textsuperscript{31} I’m grateful to Tom Hurka for pressing this point.
The virtues of restraint might, that is, function by reconfiguring our motivational structure, causing us to have fewer bad desires, rather than just by enabling us to resist the immoderate desires we do have. Even if sometimes the virtues of restraint keep us from acting badly by causing us to do better acts, or by enabling intentional omissions, they may also preclude the very development of desires that would lead to bad outcomes.

The second way to non-exhaustion involves transparent inner states. Inner states are transparent when they are manifested in involuntary outward signals, such as a blush, tone of voice, or facial gesture. Transparent states count as cases of non-exhaustion when the involuntary outward signals they involve produce good or bad consequences.\(^{32}\)

Suppose that Craig is gregarious, and this manifests itself in non-voluntary smiles and twinkles in his eyes. Craig’s smiles and twinkles are noticed by others, and cause them pleasure. Insofar as it is Craig’s cheerfulness that explains the non-voluntary smiles and twinkles, his cheerfulness produces good apart from any of the intentional acts it explains. This makes it a case of non-exhaustion.\(^{33}\)

In addition to taking pain or pleasure in the non-voluntary signals of others, we often consciously base our own actions on what we take to be the inner states of those around us, taking pains to do things because our friends are shy, our guests fussy, and our children excitable. In these situations the inner states of others seem capable of influencing our actions—by figuring as necessary components in the causal chains leading up to the actions that we (not they) perform.

Such cases can support non-exhaustion. Suppose that Larry makes a charitable donation because he is moved by Gary’s compassion. For this to support non-exhaustion, one of two things must be true. It must be true either that Larry’s generous act doesn’t depend at all on Gary’s previous generous intentional acts, in which case it will be true that Larry would have been moved by

---


\(^{33}\)I assume Craig’s cheerfulness might explain these things—involuntary expressions and the like—even though they aren’t ‘actions’ in the relevant sense. Someone who thinks they are actions in the relevant sense will not find this promising. Remarks in Pettit & Smith 2000 might be taken to support this general route to non-exhaustion, though not unambiguously.
Gary’s compassion even if Gary had never performed *any* compassionate acts. (Perhaps Larry is trying to decide whether to donate to Oxfam, and an involuntary gesture of sympathy from Gary provides the impetus he needs.) Or it must be true that Larry’s generous act depends partly on Gary’s previous generous acts, and partly on Gary’s inner generous state, as manifested in involuntary signals and the like. (Larry might see Gary donate to Oxfam, and decide to donate on this basis. As Larry is deciding how much to donate, an involuntary gesture of sympathy on the part of Gary makes Larry donate more money than he had previously planned.) In either case Gary’s generosity influences Larry by manifesting itself in non-voluntary and/or non-intentional bodily movements that do not count as actions. Gary’s generosity would under either of these conditions be a case of non-exhaustion.

Just how often we base our acts directly on the inner states of those around us depends on how transparent those around us are. It seems hard to deny that sometimes at least part of the reason we act as we do is because of the inner states of others, as those states are expressed in involuntary signals, even if it is also partly because of the manifestations of those states in intentional action. It thus seems to me that transparency gives us a second route to non-exhaustion.

The third way to non-exhaustion is perhaps the most compelling. It simply points out that intrinsically good mental states often accompany or constitute motivational states. Adams and Parfit both note that patterns of motives are correlated with quality of mental life in the ones who possess them, so that, in those who have the motives, some motivational patterns produce enjoyment or pleasure or satisfaction. Often the pleasure we take in some activity is a function of how interested and moved by it we are; we typically enjoy things we care about and are motivated to pursue more than we enjoy things for which we are motivationally apathetic.

Suppose that Beth is intensely motivated to understand and appreciate Bach, while Bryan lacks this enthusiasm. Beth may, as a result of her mindset, derive much more enjoyment than Bryan from hearing the same excellent performance of Bach. The pleasure produced here need not run through intentional acts, but can be a direct consequence or even constitutive of our motivational patterns.
As both Adams and Parfit point out, if inner states only produced pleasure via intentional acts, the best (or better) motives would always be the ones that produce the best (or better) acts. However, some motivational patterns can produce non-optimific acts but still be such that the amount of pleasure involved in being so motivated is sufficient to make these motives on-balance more valuable than other patterns of motivation that would produce only right (or better) acts.

In Parfit’s example Kate is highly motivated in her work, taking great interest and pleasure in it. As a result, she often works too hard, and occasionally suffers complete burn-outs. Even though this pattern of motives sometimes causes Kate to act wrongly (to overwork), it might be better for her to have this motivational pattern than to have one which resulted in her performing only right acts. For it could be true that the only motivational pattern that would keep Kate from overworking would be one that resulted in a much less enthusiastic and interested Kate. And it could be that the pleasure that would be lost to Kate as a result of being less interested in her work would make a bigger negative value-difference to the world than Kate’s wrong acts of overworking. The consequences of Kate’s motives are thus not exhausted by the consequences of the acts they explain.

In addition to the intrinsic good of pleasure, it is also very plausible that certain inner states are intrinsically admirable and good, and others intrinsically deplorable and bad. Unlike many views, virtue-consequentialism denies that the virtue and vice of motives is determined solely by their intrinsic value. But virtue-consequentialists should not deny that the intrinsic value of states is relevant to their virtue- or vice-status, since the intrinsic value or disvalue of states, in addition to the good or evil they more strictly speaking cause, is one way in which they make the world better or worse. A sincere desire that your friend interview well for a position that you yourself desire seems admirable and intrinsically good, in itself making the world better. Likewise, a strong desire for a rival sibling to fail in her life’s primary pursuit is intrinsically evil, in itself making the world worse. If the ‘consequences’ of inner states include, as they should, the intrinsic value or disvalue of the states themselves, the fact that inner states can be

---

intrinsically good or evil entails that the good consequences of an inner state are not exhausted by the consequences of the acts they produce.\textsuperscript{35}

If non-exhaustion is true, no matter where we set the bar for right action, and where we set the bar for virtue, it will be possible for traits to count as virtues without producing right action. Whatever the value-difference between the net value of the acts produced by an inner state and the bar for virtue, the trait can in principle make up the difference—by preventing bad action, directly causing others to act in beneficial ways, or by itself being intrinsically good enough.

Imagine, for example, that Bob is humble. Bob’s humility doesn’t produce very impressive acts—it’s always causing him to bite his tongue when it would be best for all concerned if he spoke up. When Bob does speak up, moreover, his humility ensures that he does not speak up very forcefully, and often his good ideas are drowned out by louder voices. Nonetheless, Bob’s humility also keeps him from doing and saying things he shouldn’t. Moreover, people well-acquainted with Bob benefit from his humility. They take pleasure in it, and are inspired to be less abrasive and brash in their everyday encounters with others. Finally, we can suppose that Bob’s humility is itself admirable and intrinsically good. Bob’s humility can thus make a significant positive value-difference to the world—great enough to meet a reasonable bar for virtue—despite the fact that it never issues in optimific, or indeed even significantly valuable, acts.

Given non-exhaustion, we can conclude that virtue-consequentialism does not generally ensure a strong connection between virtue and rightness. At least this is so if we assume, as I have been, an act-consequentialist criterion of rightness. Of course, virtue-consequentialists need not be act-

\textsuperscript{35} The examples above make use of the intrinsic good or evil of motives. It’s worth pointing out that for these examples to work we must individuate acts in a way that does not include the motives from which they’re done. If motives were parts of acts, their intrinsic value could be claimed to be part of the consequences of acts, in which case they would not support non-exhaustion. It seems to me more plausible that motives are explanatorily prior to acts—forming the grounds of intentions (whose good or bad effects should, I think, be included in the consequences of acts)—rather than parts of acts. But even if there is a way to deny that intrinsically good and evil motives support non-exhaustion, inner states other than motives that are clearly not part of acts (e.g., reactive attitudes) can be intrinsically good or evil; these will give us other ways to non-exhaustion.
consequentialists. Those perturbed by the conclusion that consequentialist virtues do not
guarantee consequentialist right action might respond by changing their notion of rightness.
They might, for example, say that the right act is the one recommended by the best rules, or
endorse an altogether non-consequentialist approach to right action.

The problem is that these notions of right action seem equally incapable of securing a strong
connection between rightness and consequentialist virtue. Just as, given non-exhaustion, there is
nothing to stop a state from producing lots of good consequences without causing
consequentialist right action, there is nothing to stop a state from producing good consequences
without issuing in acts that conform to optimific rules, or respect independently grounded
notions of duty. Indeed it seems to me that the only way for virtue-consequentialists to ensure a
strong connection between virtue and rightness is to identify right acts as those that proceed from
consequentialist virtues.

In the next chapter I argue that there are good general reasons for rejecting views that attempt to
explain rightness in terms of virtue. For now I will just point out that explaining right acts as
those that issue from instrumentally good states seems implausible on its face. There is nothing
in the notion of ‘issuing from an instrumentally good state’ that guarantees that acts that do so
will keep promises, respect rights, benefit others, or conform to plausible standards of rightness
generally. These things will, I hope, become even clearer below.

3.3.4. Consequentialist rightness-based virtue

If as I have argued it is always possible for states to count as consequentialist virtues without
ever producing right acts, Moore’s theory looks like a substantive departure from virtue-
consequentialism. Indeed in light of the preceding considerations Moore’s claim that inner states
must produce consequentialist right acts to count as virtuous seems sufficient to qualify his view
as at least a partly rightness-based explanation of virtue, rather than an instance of pure virtue-
consequentialism. The obvious question at this point is which type of theory we have more
reason to accept: a view like Moore’s, one that requires consequentialist right action, or a view
that places no restrictions on how good is produced. Should we care that purely goodness-based
virtue-consequentialism cannot capture a strong connection between virtue and the right?
Many consequentialists are likely to be skeptical that failing to ensure a strong connection between virtue and right action is a significant strike against a theory of virtue. After all, if the motives or traits in question really do leave the world that much better off, who cares if they never produce right action?

It seems to me, however, that many will find the claim that virtue is compatible with always doing the wrong thing dubious, or even unacceptable, on its face. How could someone who always does what’s wrong be virtuous? Moreover, how could virtue itself be responsible for consistently leading us to do wrong acts? There is, at least, something grating about these thoughts. Perhaps part of what explains the dissonance is the continuing influence of ancient and medieval approaches to virtue, which typically affirm a very strong connection between virtue and acting well. Ancient virtue theories often went so far as to claim that genuine moral excellences never issue in wrong action. If they did, they would not be truly excellent.

However, it is not just the ancients who are likely to find an extreme disconnect between virtue and rightness tough to accept. We have already seen that Moore accepted a strong connection between virtue and rightness. But we should note here Sidgwick as well, who, in his explication of common-sense or intuitional morality, assumes that virtuous conduct is dutiful or right action.36 It is Sidgwick’s assumption that virtuous action and duty are coextensive that allows him in the Methods to consider different intuitionist proposals for the grounds of rightness beneath the headings of different virtues (‘benevolence,’ ‘justice,’ etc.).37 Moreover, as I understand him, Sidgwick retains this view of the relation between virtuous action and duty throughout his assessment of utilitarianism.38 We can also, arguably, add Mill to the list of

36 “It will … involve no material deviation from usage, if we limit the term ‘Virtue’ to qualities exhibited in right conduct.” Sidgwick 1907 (1981), Bk III, ch. 2, p. 219.
37 Methods, book III.
38 This comes through in the following passage: “In order … to form a precise estimate of the extent to which Utilitarianism agrees or disagrees with Common Sense, it seems best to examine the more definite judgments of right and wrong conduct, under the particular heads represented by our common notions of virtues and duties” (Book IV, ch. 3, p. 430). The question throughout Book IV is whether or the extent to which these ‘common notions’ can be understood or
prominent utilitarians who reject pure goodness-based explanations of virtue. This can be seen in his claim that, as a sociological truth, utilitarians (presumably including himself) “resolutely refuse to consider any mental disposition as good of which the predominant tendency is to produce bad conduct.”

It is striking that Mill, Sidgwick, and Moore—arguably the three greatest utilitarians—all accepted a tight connection between virtue and rightness. Why? I think there are likely two reasons.

The first is simply that it is intuitively very plausible that virtue and right action are tightly linked. This connection is contained in our pre-theoretical notion of virtue. Common sense insists that the virtuous person is someone who at least usually does the right thing, and balks at the thought that virtue could always cause us to go wrong. It seems to me likely that Mill, Sidgwick, and Moore recognized the intuitive attractiveness of these thoughts, and simply took them for granted. What the arguments of this section show is that virtue-consequentialists cannot take them for granted, and indeed that the connection between consequentialist virtue and right action is so weak as to imply that virtue can always cause us to do what’s wrong. This gives anyone who finds it intuitively attractive that virtue and right action are closely linked reason to reject virtue-consequentialism.

The second possible reason that these utilitarian theorists, and in particular Sidgwick and Moore, may have accepted a strong relation between virtue and rightness is more complicated. In his discussion of Hume’s notion of virtue Sidgwick wonders what, for Hume, distinguishes the moral virtues from other excellences.

Why is the excellence of Virtue so strongly felt to be different in kind, not merely from the excellence of a machine or fertile field, but also from the physical beauties and aptitudes, the intellectual gifts and talents of human beings? I should answer that …

explained in terms of utilitarian principles. But to my eyes Sidgwick never questions the tight connection between virtuous action and duty or rightness.

39 Utilitarianism, ch. 2.
qualities that are, in the strictest sense of the term, Virtuous, are always such as we conceive capable of being immediately realized by voluntary effort, at least to some extent.  

Moore echoes a similar line.

There can be no doubt that Aristotle’s definition is right, in the main, so far as he says that it [virtue] is an ‘habitual disposition’ to perform certain actions: this is one of the marks by which we should distinguish a virtue from other things.  

Both Sidgwick and Moore seem to be saying that what distinguishes virtue from other things is that virtue has essentially to do with the will or action. If we accept that virtue, as it attaches to inner states, perfects the part of the soul responsible for action, we should ask how it does so.

One intuitive answer is that virtue perfects the active part of the soul by causing it to perform virtuous or excellent acts. In what does the excellence of acts consist? Prichard and Ross would insist that the excellence of acts consists solely in the excellence of the motives from which they are done. But it would, I think, also be perfectly appropriate for someone with consequentialist sympathies to conclude that the excellence of acts consists at least partly in their making the world sufficiently good: in their being objectively right. If objective rightness is a desirable feature of actions, and if virtue attaches to and perfects actions, it is very natural to conclude that virtuous action must be right action. Anyone who finds it attractive that the perfection of the will depends partly on the perfection of the acts in which it issues, and that rightness is one way in which acts are perfected, will have reason to reject virtue-consequentialism, since it cannot ensure that being virtuous involves doing what’s right.

The line of thought just sketched requires that the perfection of the will depends on more than just the will’s intrinsic properties. But presumably anyone willing to explain the virtue of inner states in terms of their consequences already accepts something like this. What this line of

---

thought does is bring out the importance of the relation between the excellence of the will and the excellence of the acts it performs. Intuitively, an excellent will should be one that performs excellent acts. If we accept this connection, it is not, as virtue-consequentialism implies, any old sort of good consequence that can suffice to make the will excellent. It is only those consequences that come about as a result of the will doing its function—acting—and doing it well—by performing acts that are right.

I’ve been offering some general reasons to reject virtue-consequentialism, construed as the view that virtue can be explained entirely in terms of instrumental goodness. Importantly, however, my defense of VA does not turn on the success of these arguments. Since VA is an account of virtuous action, all I really need is to show is that VA is more plausible than the approaches to virtuous action likely to be put forward by virtue-consequentialists. Finally, then, we should see how the preceding arguments bear on the notion of virtuous action available to virtue-consequentialists, toward the end of assessing how virtue-consequentialist accounts of virtuous action measure up to VA.

3.4. VA versus VC

Insofar as my purpose is to defend VA, I ultimately only need to show that virtue-consequentialism is less plausible than VA. This, however, raises a prima facie obstacle. VA is an account of virtuous action. I have been assuming that virtue-consequentialism is an account of what makes inner states—not acts—virtuous. So understood, virtue-consequentialism and VA might seem compatible. Indeed it is possible for VA to hold that what makes motives good is their instrumental goodness. VA would then agree with virtue-consequentialism that the virtue of motives is explained by the value of their consequences.

While it’s true that VA is compatible with a consequentialist explanation of what makes motives virtuous, VA is not compatible with virtue-consequentialism, since VA claims that the right partly explains virtuous action, whereas virtue-consequentialism is supposed to be a purely goodness-based explanation of virtue properties. If virtue-consequentialists were hospitable to explaining virtuous action partly in terms of the right, they would agree with VA that rightness is
partly explanatory of virtuous action. I, along with Mill and Sidgwick and Moore, gladly welcome such consequentialists into the rightness-based fold.

I have been assuming that virtue-consequentialism takes inner states, primarily motives or traits, as its evaluand. But it is an open question how virtue-consequentialists should move from facts about the virtue of inner states to conclusions about the virtue of acts. Perhaps the most natural strategy is for virtue-consequentialists to adopt the Prichard-Ross strategy of claiming that the virtue of acts is solely a function of goodness of motive. Since virtue-consequentialism explains the goodness of motives in terms of instrumental value, the view would then say that acts are virtuous just in case and because they express instrumentally valuable (enough) motives.

This approach returns counterintuitive conclusions about the virtue-status of particular acts. This happens in two ways.

First, virtue-consequentialism implies that intrinsically bad motives can produce acts that have good consequences, but that do not strike us as virtuous. Driver, for her part, faces up to the charge that virtue-consequentialism deems intrinsically bad inner states virtuous. She asks us to consider the Mutors, an alien race with a strong desire to beat children for pleasure when the children reach a certain age. Due to a strange law-like connection at the world in question, being beaten by the Mutors at this age causes children to live 50% longer than they otherwise would. Driver concludes that the character trait responsible for the Mutors beating the children is a virtue, since it has very good consequences. Similarly, the view of virtuous action under consideration would say that the act of beating children under these conditions is virtuous, since it is motivated by a desire that is instrumentally very good (since the consequences of the desire include the consequences of the child-beating). However, I think most people will have a difficult time accepting that beating children for fun, however instrumentally beneficial, is or can be a virtuous thing to do. I have a strong sense that the Mutors’ behavior is not virtuous. But I

---

42 Driver 2001: 56.
43 I do not think that Driver is necessarily committed to saying that the action of the Mutors is virtuous. It might be possible for her to insist that the trait is a virtue but deny that the act is virtuous, depending on her view of virtuous action and the relation between virtuous acts and virtuous traits.
do not know how to argue against Driver’s construal of the case without begging any questions. All I can do is leave it to the reader to decide which judgment about the case is more plausible.

Second, the view returns counterintuitive results in cases where instrumentally good motives produce acts with bad consequences that do not strike us as virtuous. Such motives— instrumentally good states that cause acts with bad consequence—are made possible by non-exhaustion. Imagine that Tim loves attempting to kick people in the shins. It’s not that he enjoys the actual kicking, when he succeeds; indeed this causes him very little or no pleasure. What causes Tim enjoyment is the pursuit of shin-kicking: different strategies for aiming and striking the blow; different footwear for administering different types of kicks; strategies for sneaking up on people and kicking them when they least expect it; and so on. Tim loves all this so much so that the giddiness he gets from pursuing his program of shin-kicking outweighs the suffering involved in the actual kicking. If Tim takes enough pleasure in shin-kicking, it will, on the view in question, be virtuous for Tim to kick shins, since it is the motivational pattern involved in kicking shins that brings Tim so much joy. But Tim’s acts of shin-kicking do not seem virtuous, no matter how much enjoyment Tim takes in them.\(^{44}\)

We should see that VA is not subject to these difficulties, even without taking a stand on the truth of act-consequentialism. Since VA ultimately explains good motivation as admirable motivation (see chapter five), and takes admirable motives to be necessary for virtuous action, it precludes the first type of case, since intrinsically bad motives are not admirable. And since VA requires acts to be objectively right, it also blocks the second type of case, at least when acts like Tim’s are objectively wrong.

Of course, virtue-consequentialists might try to avoid the counterexamples just pressed with a different approach to virtuous action. They could, for example, make virtuous action a more complex function of both the instrumental goodness of an act’s motive and the instrumental goodness of the act itself, considered apart from its motive. On this view acts would be virtuous

\(^{44}\) Cases like Tim’s show how implausible it would be to identify the right act as whatever proceeds from instrumentally good motives; such views would deem Tim’s acts of shin-kicking right.
just in case and because they are done from instrumentally valuable (enough) motives and themselves have good (enough) consequences. However, while this takes care of cases like Tim’s, it does not succeed at turning back the first type of counterexample: the view would still deem it virtuous for the Mutors to beat children. Indeed it seems to me that any virtue-consequentialist approach to virtuous action will be subject to counterexamples of the general Mutor type.

Just as significant, to my eyes, is the fact that virtue-consequentialist theories of virtuous action cannot, without making rightness partly explanatory of virtuous action, capture the organic unity approach to virtuous action. We can put the argument for this in terms of a dilemma. Either virtue-consequentialist theories of virtuous action require virtuous acts to be objectively right, or they do not. If they do, the best explanation for this is that rightness explains virtuous action (rather than virtue explaining rightness; see chapter four for argument). In that case virtue-consequentialist accounts of virtuous action would be at least partly rightness-based, and proponents of VA would have cause to celebrate.

If on the other hand virtue-consequentialists do not require virtuous acts to be objectively right, they will presumably toe the Prichard-Ross line, saying that acts are virtuous when and because they are done from virtuous motives. But if non-exhaustion is true, virtuous motives, construed as instrumentally good motives, need never produce right action. So virtuous acts need never, on this approach, be objectively right. It follows that virtuous acts need not constitute organic wholes, comprised of admirable motives + objective rightness, with their own distinctive value. Virtue-consequentialism will thus fail to capture an important way in which human acts can be particularly good instances of their kind.

Moreover, while virtue-consequentialism will have no trouble explaining the social benefits of virtue, it will, by failing to capture these unities, fail to explain the individual benefits of these unities—the way in which these wholes make the lives of those who instantiate them better—beneath the concept of virtue. Virtue-consequentialist approaches to virtuous action will thus be unable to unify and shed light on the connections between virtue and benefit in the way VA claims to accomplish (sec. 1.6.).
All in all, then, it seems to me that we have strong reason to reject virtue-consequentialism generally, as well as strong reason to prefer VA to virtue-consequentialist theories of virtuous action. These reasons stem most basically from the difficulty of stating virtue-consequentialism in a way that does not entail unacceptable conclusions about virtue and vice, and the fact that virtue-consequentialism cannot capture a strong connection between virtue and right action. Whether or not these difficulties warrant us to reject virtue-consequentialism outright, they should at least lower our credence that it is true, and raise our credence that VA is preferable to virtue-consequentialist theories of virtuous action.

3.5. Recap
In the first two chapters I argued that we have good reason to prefer VA both to the Prichard-Ross approach and to subjective rightness-based explanations of virtuous action. In this chapter I’ve argued that a third prominent alternative to VA, virtue-consequentialism, suffers serious shortcomings. While there are other possible theories of virtue inconsistent with VA, I take these three—the Prichard-Ross approach, subjective rightness-based theories, and virtue-consequentialism—to constitute the most promising alternatives to views that posit a tight connection between virtue and objective rightness. Given this, if the arguments of the first three chapters work, we have strong reason to accept a strong logical connection between virtue and objective rightness.

However, this does not yet warrant us to accept clause (ii) of VA. Clause (ii) claims more than that acts are virtuous only if they are right. It also claims that part of the reason why acts are virtuous is because they are right. And this, the explanatory claim, does not follow from the claim that rightness is logically necessary for virtue. In particular, it’s compatible with virtuous acts being of necessity right that virtue explains rightness—that what makes acts right is the fact that they are virtuous—rather than vice-versa. The next chapter is devoted to this possibility.
Chapter four

Against virtue-based rightness

4.1. Entailment and explanation

I take the arguments of the first three chapters to provide us with good reason to accept that acts are virtuous only if they are right. Even if these reasons were decisive, they would not warrant us to accept VA, since VA claims that the virtue of acts depends partly on their rightness, which doesn’t follow from the claim that acts are virtuous only if they are right. Many will, moreover, think that there is a way to capture a necessary logical connection between virtue and rightness without making virtue depend on rightness. This involves simply switching the order of explanation, so that an act’s being right depends on it being virtuous, rather than vice-versa.

This sort of project—grounding the rightness of acts in facts about virtue—describes a spate of recently popular ‘virtue ethics.’ Because VA claims that rightness explains virtue, it implies that projects like this are misguided. Rightness cannot explain virtue, as VA claims, if virtue explains rightness. This means that we should accept VA only if the case for concluding that rightness explains virtue is stronger than the case for concluding that virtue explains rightness.

We should be clear about distinguishing two theses at the start. The first is that acts are right just in case they are virtuous. The second is the stronger claim that acts are right because they are virtuous. Since the second thesis entails the first, there are at least two ways to argue against the second thesis.

First, we can show that the second thesis is false by showing that the first thesis is false, and in particular by marshalling examples of right acts that aren’t virtuous. But we can also, secondly, employ arguments directly against the claim that what makes acts right is the fact that they are virtuous. We can try to show that, even when acts are both virtuous and right, facts about their virtue are not the best way to explain their rightness. I will employ each of these strategies in what follows.
I’ll consider three prominent virtue-based theories of rightness: the virtue-subjunctive approach pioneered by Rosalind Hursthouse; the targeted view defended by Christine Swanton; and the agent-based theory championed by Michael Slote. I will argue that each of these approaches faces significant difficulties, which, in conjunction with the independent attractiveness of VA, warrant us to prefer VA to virtue-based rightness.

4.2. *Virtue-subjunctivism*

Perhaps the most familiar virtue-based theory of right action is the one pioneered by Rosalind Hursthouse.¹ According to Hursthouse an act is right just in case it’s an act a virtuous agent would characteristically do in the circumstances. Because Hursthouse sees her view as an alternative to consequentialism and deontology, she is, I think, best read as endorsing the stronger explanatory claim as well: as saying that *what makes* acts right is the fact that they would be done by virtuous agents. I’ll call views of this sort ‘virtue-subjunctivism.’²

If the right act is the act the virtuous person would do, how should we understand the virtuous person? Virtue-subjunctive theories say that the virtuous person is the one who has and exercises the virtues, and that the virtues, in turn, are to be understood at some deeper level—as those traits which possess some common feature which marks them as virtues. Hursthouse’s preferred view is that the virtues are those character traits that are required for and contribute to the *eudaimonia* or flourishing of the agent. But Hursthouse recognizes that her theory of right action is compatible with understanding the virtues differently, so long as the concept of right action is not employed in the explanation of virtue.

---

¹ Hursthouse 1999.
² There are different ways to understand ‘the act the virtuous person would do.’ One important question is whether ‘the act the virtuous person would do’ should be understood to include the motives a virtuous person would characteristically be moved by—to be a virtuously motivated act—or rather in abstraction from the motives that explain it. Since virtue-subjunctivisms on which ‘the act a virtuous person would do’ is equivalent to ‘a virtuously motivated act’ reduce to versions of agent-basing, my criticisms of agent-basing below apply to them. In this section I’ll focus instead on versions of virtue-subjunctivism on which doing ‘the act a virtuous person would do’ involves more than just acting from good inner states.
Much of the criticism of virtue-subjunctivism has focused on the extensional adequacy of the view. This is a version of the first critical strategy mentioned above. These critiques typically consist in advancing examples of acts that are right but would not be done by the virtuous, thereby undermining the putative logical entailment from ‘being an act the virtuous person would do’ to ‘being right.’ I won’t be concerned with this debate here. One reason for this is that it may be compatible with VA to hold that acts are right only if the virtuous person would do them. If it turns out that a virtuous person would do an act just in case the act is virtuous, VA will agree with virtue-subjunctivism that acts are right just in case the virtuous person would do them.

What is not compatible with VA is virtue-subjunctivism’s further claim that what makes acts right is facts about virtue. VA claims that the explanatory connection runs the other way around. I will aim my criticisms of virtue-subjunctivism at the explanatory claim that acts are made right by the hypothetical responses of virtuous agents. This explanatory claim is, to my eyes, the more interesting aspiration of virtue-subjunctivism, even if it is not as frequently discussed in the literature.

Virtue-subjunctivism claims that hypothetical facts about the responses of virtuous agents explain why acts are right. It seems to me that the most basic and damaging objection to these views is that it is more plausible to explain the rightness of acts in terms of the features of the world to which virtuous agents respond, without needing to appeal to the responses of virtuous agents.

To start note that it seems perfectly appropriate to ask why the virtuous person responds one way rather than another from situation to situation. And it is, likewise, extremely natural to list in response the features of situations in which virtuous people find themselves—features that call for or demand certain responses. But then appeal to hypothetical facts about the responses of virtuous agents begins to look otiose. Why can’t we explain the rightness of acts just in terms of

3 Johnson 2003. For recent attempts at rebuttals see Tiberius 2006 and McAleer 2010.
4 See Driver 2006. Kawall 2009 has a response, which I take up below.
the features of context that demand the response—the features of the situation on which the appropriateness of the response supervenes—rather than the response itself?

Suppose you encounter a blind person who is obviously struggling to navigate a busy intersection.\(^5\) You are in a rush to catch a bus out of town—so rushed that if you stop to help you may very well miss your bus, thereby setting you back several hours. What would the virtuous person do? Suppose she would stop to help. Why would the virtuous person stop to help? Plausibly, it is because the blind person is struggling and needs help, and there are no more important claims on her. But if this fact—that a person is struggling and needs help—explains what the virtuous person would do in this context, why doesn’t it explain what you should do in your context? The reason it’s right to help is that someone in front of you is struggling and needs help, and you can help without sacrifice anything of comparable or significant importance. The fact that the virtuous person would help, while true, is explanatorily idle.

To meet this objection virtue-subjunctivists must provide compelling grounds for denying that the features of context to which virtuous agents respond themselves have moral significance, apart from facts about the responses of virtuous agents. The objection assumes that the virtuous person responds to features of context that provide reasons to act one way rather than another. But perhaps this is not so. Instead, that certain features of context have moral weight might depend on it being true that the virtuous person would respond to them in relevant ways. If the virtuous person wouldn’t respond to these features in these ways, it wouldn’t be true that these features have evaluative, and in particular right-making, significance.\(^6\)

One way to develop this thought would be to invoke particularism, claiming that one and the same consideration can have reason-giving force in one context but not another. If what makes helping the blind person right is that it’s a case of helping someone in need (rather than because it’s what a generous person would do), helping someone in need would always be right. But it’s not. Helping two shady-looking characters struggling to get a bronze statue down the museum

\(^5\) See Ross, *Foundations* p. 168

\(^6\) While Setiya 2001 might look like a defense of this view, he explicitly stops short of claiming that what *makes* something a reason for action is facts about virtue (rather than that there is an equivalence between reasons and facts about virtue).
stairs at 3am, for example, is not right (assuming that these men are stealing the statue solely to pad their coffers, etc). If helping is not always good or right-making, but sometimes is, what explains why it is right-making in some cases but not others? The view under consideration claims that the best explanation is the hypothetical responses of virtuous agents. The reason helping the blind is person right, and helping the statue thieves is not, is because a virtuous person would help the blind person but not the thieves.

This, however, merely pushes the problem back a step. Even if particularism is true, it’s still appropriate to seek an explanation for why the virtuous person helps the blind person but not the shady-looking men. It’s plausible that, on particularism, what makes helping right-making in one context but not the other is the presence of further features of context. What explains why it’s right to help the blind person but not the shady-looking men has to do with the fact that the blind person is merely trying, and struggling, to get across the street, while the shady-looking men are taking something that does not belong to them, in order to pad their coffers.

We can press the appropriateness of asking for an explanation for virtuous responses by asking whether virtue-subjunctivists want to affirm that in identical situations ideally virtuous persons of necessity respond the same. Since denying this would land virtue-subjunctivism at odds with the widely accepted belief that the moral supervenes on the natural, presumably virtue-subjunctivists will accept it. If they do, however, and if the appropriateness of virtuous responses isn’t explained by features of context, we should ask why virtuous agents in identical situations would of necessity respond the same. Is this just a coincidence? If not, there must be something that explains it. The only candidate seems to be features of the relevant context. But then these features seem to have moral importance prior to the responses themselves.

If on the other hand we deny that these features in themselves demand or call for the relevant responses, we seem to have no explanation for why identical situations elicit identical responses from virtuous agents. Worse yet, if the responses of virtuous agents are not based on features of

---

7 Herman 1981.
8 I owe this way of developing the response to Sergio Tenenbaum.
9 On the assumption that facts about how ideally virtuous agents would respond count as ‘moral’ facts.
context that themselves call for and make appropriate certain responses, the responses of the virtuous seem capricious and ungrounded. Why exactly should we care about how the virtuous person would respond, if she is not responding to things of value and importance in their own right?

There is something very much like the Euthyphro dilemma here. Divine command theorists hold that God’s commands determine rightness, imbuing acts with moral significance they wouldn’t have otherwise had. But why should we care about God’s commands, if they aren’t responses to real values? Why should we think God’s commands are binding for us, if they’re grounded merely in the caprice of God’s will? If on the other hand God’s commands are responses to real value, we do not need them, over and above the values on which they are based, to ground rightness. Likewise, if the virtuous person does not respond to features of situations with intrinsic evaluative force—if certain features of the world do not on their own demand a response from her—why care about these responses, or think that they are constitutive of rightness?

It might be replied that we should care about hypothetical virtuous responses precisely because they proceed from virtue. Consider an analogous reply to the Euthyphro problem. Theists can insist that we should care about what issues from the divine will because God is of necessity morally perfect. Indeed on common theistic premises God is identical to or the source of the good. Likewise, on the assumption that the hypothetical responses of the virtuous proceed from virtue, we have reason to care about them and consider them morally binding.

It is ultimately difficult to know whether divine commands not grounded in values external to God could be morally binding for us. It seems even more difficult to know whether we have sufficient reason to care about responses from hypothetical idealized virtuous agents. One difference between divine command theories and (non-theistic) virtue-subjunctivisms is that theists take God actually to exist, and the workings of God’s will to be (in some sense) concrete and real. Virtue-subjunctivists presumably do not claim the same for hypothetical idealized virtuous agents. This seems to make the authority of idealized virtuous responses even harder to understand. We at least want to know more about the hypothetical virtue from which the
relevant responses issue. Are we supposed to imagine hypothetical virtuous agents as on a par with God, morally perfect and omniscient, unable to do or get things wrong? What exactly about their virtue explains this?

While all explanations must stop somewhere, brute facts about hypothetical responses from idealized virtuous agents do not seem the best kind of facts to constitute rightness. It seems far simpler and more intuitive to hold that facts about the situations in which we find ourselves—facts about how our acts bear on pain and pleasure, friendships, promises made, achievement, and so on—have their own moral significance than to think that these things are important because of true counterfactuals about virtuous responses to them.

We should, however, consider one virtue-subjunctivist attempt to deal with these difficulties, which I suspect would be characteristic of many who find virtue-based approaches to ethics appealing. Jason Kawall has argued that we can avoid the basic objection—that to avoid being capricious the responses of virtuous agents must be based on some more basic moral consideration—by appeal to a dispositional theory of value. Kawall points to an analogy between color-perception and the responses of virtuous agents.

It is considered quite plausible among many that color properties are not properly understood entirely independently of the reactions of observers. For example, if we consider objects that are yellow, it seems highly unlikely that we would group just these objects together were it not for the reactions of normal observers under normal circumstances. The same would be true with other colors. We “carve up” the world in a way that would seem rather arbitrary were it not for the existence of human observers with common, shared reactions to these objects. Yet this is not arbitrary. Given our human visual systems, the world will affect us in common, predictable ways (and we cannot simply choose to see things otherwise).

---

10 And those influenced by McDowell in particular.
Kawall goes on to suggest that the same may be true of the responses of virtuous agents. Given the virtuous person’s psychology, she will respond to certain considerations in certain ways. But just as it would be misguided to conclude that the color green exists independently of human visual responses, so is it a mistake to conclude that the considerations to which the virtuous respond have evaluative significance apart from their responses.

This way of responding to the objection raises deep issues, which are well beyond the scope of this chapter. One thing to say is that Kawall’s reply does not make virtuous responses on their own explanatory of rightness, since on this view the considerations to which virtuous agents respond have some kind of evaluative significance—whatever it is that make these considerations, rather than others, such as to elicit particular responses from the virtuous (similarly to the way in which facts about wavelengths constrain color properties).

Perhaps more importantly, appeal to a dispositional theory of value—constituting as it does a controversial metaethical and metaphysical picture about nothing less than the relation between world, mind, and value—seems a lot for defenders of virtue-subjunctivism to take on. Even if a dispositional theory of value can help rebut Euthyphro-style objections to virtue-subjunctivism, this will not by itself warrant us to accept it. Moreover, with respect to the choice between VA and virtue-subjunctivism, many will take it to be a cost of virtue-subjunctivism if it turns out to require a dispositional theory of value. It would then be true that VA is, and virtue-subjunctivism is not, compatible with a wide range of views about the ultimate nature and explanation of moral properties. While dispositional theories may turn out correct, they may not. Anyone not already convinced of the truth of a dispositional theory of value should conclude that Kawall’s rejoinder places virtue-subjunctivism at a disadvantage relative to VA.

Finally, I argued in the Introduction that action is a genuine bearer of virtue, and that acts need not proceed from stable virtues to be virtuous. If this is right, it would be strange to find an explanatory connection between virtuous traits and right action that did not first run through
virtuous acts.\footnote{It would be strange even if we do not think that virtuous traits are to be analyzed in terms of virtuous acts. So long as virtuous acts are a basic bearer of virtue, they are the obvious place to look for a connection to right action.} This suggests that virtue-subjunctive views and any other approach that tries to explain right action directly in terms of a connection to the \textit{virtuous person}, rather than directly in terms of a connection to \textit{virtuous action}, is on the wrong track.\footnote{I take this to include consequentialist views that explain the virtues as those traits productive of (enough) good, and then explain right action as those acts that express or are explained by these traits. Adams 1976. Driver considers and rejects views of this sort (2001: 71).} If actions are fundamental bearers of virtue, the most promising route to a virtue-based theory of right action is directly through virtuous action.

4.3. \textit{Virtuous action-based approaches}

Among virtue-based theories that take virtuous action as their main focus, two strategies stand out.\footnote{Cf. Thomson 1997.} Both begin by specifying an account of virtuous action and then attempt to explain rightness in terms of virtuous action: acts are right either because they are virtuous, or because they are not vicious. The first type of view makes the virtue of acts depend on properties of acts other than, or in addition to, the value of the motives from which they are done. Such acts must, perhaps in addition to being well-motivated, also measure up to extra-mental standards: must ‘accord with’ or ‘hit the target’ of virtue, where hitting the target is not reducible to being virtuously motivated. Call these ‘targeted’ virtue-based views. The second type of view makes virtuous action a function of motives. An act is virtuous, on this view, when and because it expresses occurrent virtuous motivation, and vicious when and because it expresses vicious motives. Call these (following tradition) ‘agent-based’ views. I’ll begin with targeted views.

4.3.1. \textit{Targeted theories}

Targeted theories claim that acts are virtuous just in case and because they ‘accord with’ or ‘hit the target’ of virtue, and that acts are right just in case and because they are virtuous. To keep targeted views distinct from agent-basing, ‘hitting the target’ of virtue must involve more than expressing good inner states, though expressing good motives might be \textit{one} of the conditions of hitting the target. Because targeted views can make ‘hitting the target’ depend on both these things—being well-motivated and actually realizing certain ends—it’s possible for targeted
views to agree with VA that acts are virtuous just in case they are well-motivated permissible successes. But there would still be dispute between VA and targeted views, over the status of the aims realized; and in particular over whether it is independently right to realize the ends in question, or whether the rightness of doing so depends on facts about virtue. VA claims that the ends in question are independently permissible or right to realize, while targeted views make their rightness depend on being the target of virtue.

To filter out noise, I’ll limit the discussion at this point to targeted views on which hitting the target of virtue depends just on realizing certain ends, and not at all on acting from good motives. The objections against agent-basing in the next section will apply to any targeted view that makes acting from good motives either necessary or sufficient (or both) for hitting the target.

The basic challenge faced by all targeted theories is to specify the ‘hitting the target’ metaphor in a way that renders the theory plausible. Since Christine Swanton has done the most to develop targeted theories, I’ll focus on her version of the view.\(^{15}\)

Swanton claims that an act is virtuous in some respect V (courageous, generous etc.) just in case it hits the target of V (courage, generosity), and that an act is right just in case it is overall virtuous.\(^{16}\) She then explicates ‘hitting the target’ of virtue as follows.

\[
\text{Hitting the target: Hitting the target of virtue is a form of success in the moral acknowledgement of or responsiveness to items in its field or fields, appropriate to the aim of the virtue in a given context.}\(^{17}\)
\]

\(^{15}\) I think something like the basic view here is also probably held by Thomson 1997, though unfortunately she doesn’t present her view of virtuous or vicious action in enough detail to be able to say for sure.

\(^{16}\) I’ll leave aside the possibility that some acts that hit the target of a virtue might be less good than other acts that hit the target of a different virtue (or virtues), as well as other complications to the view, since nothing below turns on them. See Swanton 2003 (pp. 234-244) for the complications. Swanton also never claims that what makes acts right is the fact that they are virtuous; she just states the relation between right and virtuous action as an equivalence. I assume, however, that she intends her view as a virtue-based theory of rightness in the relevant sense, given the way she contrasts her theory with the views of Slote and Hursthouse.

\(^{17}\) Swanton 2003: 233.
The idea seems to be that virtues make their possessors responsive to different types of considerations, in different types of contexts. Courage makes us properly responsive to situations where danger and fear are present; justice makes us properly responsive to situations involving the distribution of goods and evils; benevolence makes us properly responsive to situations where well-being and need are at stake; and so on. To hit the target of a virtue is to succeed at responding to the situation governed by the virtue in question. Hitting the target of courage thus might involve standing one’s ground (literally) for a just cause; and I might hit the target of temperance by not having a third helping of cake.

Since we are considering views on which hitting the target does not involve having good inner states, the motivational states of the agents in these examples are irrelevant to hitting the target. So long as I actually stand my ground, or refuse the third piece of cake, I hit the target—even if I am conflicted and do so with difficulty. This feature of targeted views is attractive, insofar as it allows targeted views to capture the intuitive importance of outcomes for determining rightness. Realizing certain outcomes, not merely trying to realize them, is necessary for hitting the targets of different virtues.

So we get an explanation of ‘hitting the target’ in terms of realizing the characteristic aims of particular virtues. This doesn’t take us to all-things-considered rightness, as Swanton realizes, since some acts that hit the target of particular virtues are wrong. If the target of benevolence is helping those in need, I hit the target whenever I succeed at helping. But of course it isn’t always right to help those in need—it usually isn’t right to help looters who are having a hard time breaking shop windows, or the shady-looking men on their way out of the museum. Targeted views face a choice at this point. They can either look for a function to take us from facts about hitting the target of particular virtues to conclusions about overall or all-things-considered virtuousness, which could then be used to explain all-things-considered rightness. Or they could rest content with having provided a virtue-based explanation of right-making factors or pro tanto rightness. If the attempt to explain pro tanto rightness in terms of hitting the target of particular

---

18 This resembles Nussbaum’s 1988 Aristotelian view of how to individuate virtues.
virtues fails, the more ambitious version fails as well. So I’ll proceed by arguing against the more modest version of the view.  

The basic problem with targeted views is that, no less than virtue-subjunctivism, they make otiose considerations about virtue the ultimate determinants of rightness. Suppose that hitting the target of benevolence consists in helping those in need. Suppose, further, that you perform an act that fits this description—you acquire basic medical training and spend time aiding people in a famine-stricken area—and that in doing so you perform an act that’s pro tanto right. When we wonder why helping here is pro tanto right, we think first of the people who are helped. Targeted views claim that what makes it the case that helping the suffering is morally significant is the fact that doing so hits the target of benevolence. But this looks like the wrong explanation. Helping the suffering is right simply because they are suffering.

It is likely to be replied that suffering itself would not have moral significance were it not for the fact that certain virtues (benevolence, kindness, compassion, etc.) are sensitive to suffering and aim at its alleviation. But do we really need an explanation for why suffering is morally significant? And even if we do, the fact that suffering relates to such-and-such virtues in certain ways is, on its face, an implausible candidate for an ultimate explanation for why suffering is bad.

To press this point further, consider E1, which is what targeted views claim, and E2, which switches the order of explanation in a way that’s consonant with VA.

E1: Alleviating suffering is morally important because it is the proper aim of a virtue.
E2: Alleviating suffering is the proper aim of a virtue because it is morally important.

Because explanation is asymmetrical, E1 and E2 are incompatible. The moral importance of $p$ cannot be explained by the fact that $p$ is the target of a virtue if the fact that $p$ is the target of a

---

19 It’s worth noting a distinct objection from the one I pursue below. For targeted views to be successful, they must hold that, for every type of pro tanto right-making feature, there is a virtue that is sensitive to that type of feature. For argument to the contrary, Sreenivasan 2009.
virtue is explained by the fact that \( p \) is morally important. Because \( E_1 \) is true just in case \( E_2 \) is false, and vice-versa, our credence in \( E_1 \) should be inversely proportional to our credence in \( E_2 \), and vice-versa. For targeted views to be compelling, then, the case for \( E_1 \) must be stronger than the case for \( E_2 \).

On its face, however, \( E_1 \) seems less promising than \( E_2 \). It seems far simpler to explain the fact that alleviating suffering is the target of (say) kindness by appeal to the importance of alleviating suffering than to explain the importance of alleviating suffering by appeal to the fact that it is the target of kindness.

In large part the difficulty with \( P_1 \) stems from the fact that, as with virtue-subjunctivism, it seems perfectly reasonable to ask why certain ends are the aims of certain virtues. Why should we think that benevolence or kindness or compassion aim to alleviate suffering? Surely the most natural answer involves the thought that suffering is bad. Since targeted views deny that alleviating suffering has evaluative significance in itself, apart from the fact that it is the appropriate end of virtues, they cannot say this. But then targeted views, like virtue-subjunctivism, cannot meet this seemingly appropriate demand for explanation. Targeted views cannot explain why benevolence or kindness responds to suffering and aims to alleviate it, in any way that can satisfy our natural desire to make deeper sense of the normativity of these responses. We are left wondering exactly why responding in these ways is something we have reason to care about and do.

\( P_2 \), on the other hand, can meet the demand for explanation: it can fully affirm that kindness aims to alleviate suffering at least in part because suffering is bad. At the same time, \( P_2 \) is compatible with any number of deeper explanations for the badness of suffering, should we feel that one is needed. This makes \( P_2 \) more attractive—indeed I think far more attractive—than \( P_1 \). On this basis I think many will see here decisive reason to reject targeted approaches.

4.3.2. Agent-basing

The final virtue-based theory of rightness is agent-basing. As I’ll understand them, agent-based views begin with the claim that acts are virtuous when and because they express occurrent
virtuous motivation, and vicious when and because they express vicious motivation. From here they claim either that acts are right when and because they are virtuous; or wrong when and because they are vicious, and right when they are not wrong. Different agent-based theorists can, within this basic framework, disagree over the details of the view. Michael Slote, who has led the way here, prefers to make the value or admirableness of right-making motives primitive, and limits the class of right-making motives to one: universal benevolence. But it’s possible to disagree with Slote on one or both of these points while still counting as an agent-based theorist, in the sense that interests me.

Up till now I have utilized the second argumentative strategy for arguing against virtue-based theories of rightness (see earlier, sec. 4.1), arguing directly against the idea that virtue explains rightness without calling into question the extensional equivalence of right acts and virtuous acts (or acts the virtuous would do). In this section I will switch gears and utilize the first argumentative strategy, arguing that certain well-motivated acts are not in fact right. If some acts that are well-motivated (in the relevant sense, whatever it is) are not right, facts about good motivation cannot provide a general explanation for rightness, in the sense we’re interested in.

The most basic objection to agent-basing is that it cannot capture the distinction between evaluating the deontic status of acts and evaluating the goodness of the motives from which they are performed. As we’ve seen (ch. 1), this distinction finds able proponents in Prichard and Ross. According to them, evaluating the rightness or wrongness of acts is one thing; evaluating the goodness or badness of motives another.

Prichard and Ross argued for this primarily on the grounds that making rightness depend on motives results in violations of ‘ought’ implies ‘can.’ If I’m obligated to act from a good motive, I must be able to act from a good motive. But in some situations good motives may be unavailable to me. Suppose I am unable to satisfy my teaching duties from anything other than worry about what my students will say about me on end-of-term evaluations, which I know are assiduously studied by my chair. In that case the only motive I can teach from is the motive of self-interest. This doesn’t seem like an admirable motive. If it’s not, I will not be able to teach from a good motive. But then what seemed to be a professional obligation of mine in fact is not.
If it is true that ‘ought’ implies ‘can,’ then, if I cannot teach from good motives, I am, on agent-basing, not obligated to teach at all. That’s clearly wrong. The fact that I cannot teach from better motives in the situation does not make teaching any less of a duty for me.

Objections from ‘ought’ implies ‘can’ are common criticisms of contemporary agent-basing.\textsuperscript{20} However, they will not by themselves be persuasive against all agent-based theories, since some versions of agent-basing do not run afoul of ‘ought implies can.’ If for example an agent-based view says that an act is right just in case and because it expresses the best motive it’s possible for the agent to express in the situation, we will always be able to do what’s right.\textsuperscript{21} We cannot cast doubt on these versions of agent-basing on the grounds that they run afoul of ‘ought’ implies ‘can.’

But we can, I think, recognize an even more basic difficulty that comes from abandoning the distinction between motive and rightness. Any theory that makes all-things-considered rightness depend entirely on acting from good motives is stuck with unacceptable verdicts about the deontic status of acts, even apart from considerations about ‘ought’ implies ‘can.’ Many will, quite simply, want to acknowledge cases where agents do the right thing from bad motives. Consider a man who dives into a swimming pool to save a drowning child for the sole purpose of impressing the child’s mother, so that he might sleep with her.\textsuperscript{22} Saving the child’s the right thing to do; it’s just done for the wrong reason. Agent-basing must, however, say that, insofar as the man saves the child from bad motives, saving the child is wrong.

In response to cases like this, Slote seems to deny that the act of saving the child is intuitively right. There’s nothing counterintuitive, he insists, about calling the life-saving act wrong.\textsuperscript{23}

\textsuperscript{20} Hurka 2001; Jacobson 2002. Some try to defend Slote here by denying that his theory intends to yield conclusions about what we \textit{ought} to do. It is only meant to yield conclusions about what is in fact right (van Zyl 2005; Russell 2009). I don’t understand this strategy or see how it helps.

\textsuperscript{21} This is essentially the view of the 19\textsuperscript{th} century British philosopher James Martineau. More recently, see Doviak 2011.

\textsuperscript{22} The example is from Das 2003.

\textsuperscript{23} This is apparent in Slote’s concession that the prosecutor who prosecutes from malice (in Sidgwick’s example) does what is wrong (Slote 2001: 15).
What is counterintuitive, says Slote, is that the man has no duty to save the child. But, Slote points out, agent-basing isn't committed to this. For consider what motives the man would express by not saving the child. They would be bad, exhibiting an insufficient desire to help. Since the man has a duty to refrain from acting on bad motives, he has a duty to save the child. Slote thinks this general strategy yields a distinction between doing one’s duty for the right reason (saving the child out of a desire to help him), and doing it for the wrong reasons (saving the child out of the desire to sleep with the mother). It’s just that in the latter case the act is wrong. Slote concludes that, while saving the child to sleep with the woman is wrong, the man nonetheless has a duty to save the child.

For my part it seems clearly false that saving the child to sleep with the woman is wrong. It sounds much more plausible to my ears to say that saving the child is right, but that the motive behind it is bad. This said, there is likely to be disagreement over whether saving the child to sleep with the mother is wrong. Certain Kantians, for example, are likely to want to include the man’s reasons for action as part of the description of the act itself, in which case it may be less plausible to insist that it is right.

We may be able to make more progress by considering the converse type of case: one involving a well-motivated agent who does what’s intuitively wrong. Consider an agent who must choose between doing an intuitively right act from lackluster or bad motives, and doing an intuitively wrong act from good motives. Suppose that I am scheduled to work this afternoon, and that my co-workers are very much depending on me to show up. Alas, the only motive I have for showing up, the only motive that will succeed at getting me to work, is the motive of material profit. I don’t care that my co-workers are depending on me; I simply want the money. Suppose, moreover, that I have been raised to treat my job, and work generally, as a mercenary affair, in a way that renders me non-culpable for these beliefs. I have been raised to think that my only reason for going to work, and for satisfying the professional obligations I incur as a result of taking a job, is the money. If I am willing to forego the cash, there is, in my mind, nothing whatsoever wrong with me bailing.

Now imagine that on my way to work one day I receive a call from a friend who asks me to help him move across town, that afternoon. I find myself swayed to help by the motive of generosity. Since I am scheduled to work, however, I must choose. I can either go to work from the motive of profit, or help from the motive of generosity. Moreover, if I choose to help my friend, I will, in my own mind, be doing nothing wrong by bailing on work. The question in my mind is just whether it is worth passing up the money I will earn from work in order to help my friend. According to agent-basing, the right act in this situation is the one that expresses an admirable motive. The motive of generosity seems admirable, while the motive of profit does not. Moreover, because I do not think I am obligated to go to work, acting on the motive of generosity will not involve acting against my judgment about what is right, in a way that would detract from the goodness of my motives. On this way of filling in the case my overall motivation in helping my friend—the motive of generosity, in conjunction with a willingness to forego the cash from work; plus (non-culpably) lacking a belief that bailing on work is wrong—seems clearly good. On the other hand, my overall motivation in going to work—the motive of profit, in conjunction with the belief that helping my friend is not worth the money I will lose by missing work—seems lackluster, even slightly bad. According to agent-basing, then, the right act is for me to skip work and help my friend. This is likely to strike many of us as false. The right thing for me to do is go to work. It would, in this situation, be wrong for me to help my friend, given that my co-workers are depending on me. Agent-basing returns the wrong verdict about the case.

We should note that the objection above does not appeal to the intuitive relevance of outcomes for determining rightness. Regardless of whether or not I actually succeed at helping my friend, the choice to help my friend rather than go to work seems wrong. Those who think that outcomes matter for determining rightness can mount a further and even more compelling criticism of Slote’s view.

Consider Magoo, a well-intentioned bumbler who leaves trails of destruction in his wake. Agent-basing says that, insofar as Magoo is well-motivated, his acts are right—regardless of how much devastation they produce. This is hard to swallow. Or imagine a case where I have been bitten by a deadly poisonous snake, and it is up to you to give me an antidote. Our wilderness
guide has deserted us, and you do not know which of the two vials marked ‘antidote’ in her bag is the right one. Suppose that, gravely concerned for my well-being and hoping to save me, you give me the antidote in one of the vials. Unfortunately, you choose the wrong vial—the antidote I need is in the other one—and I die. Agent-basing entails that, so long as your motives in giving me the vial are good, you have done what is right. But, no matter how good your motives, and even if you are not to blame for it, there is a clear sense in which you have made the wrong choice. You should have given me the other vial.

What emerges from examples like this is that agent-basing’s singular focus on the internal states of agents makes it impossible for the theory to capture the relevance of what actually goes on in the world, outside the agent’s mind, for rightness. Slote acknowledges that this is a legitimate concern, but tries to deal with it by claiming that the genuinely well-motivated person will take steps to discern the likely consequences of her actions, and that her choices will reflect this.\textsuperscript{25} If Magoo were genuinely well-motivated, he would have taken sufficient steps to ensure that his acts would not result in devastation. But this cannot defeat all forms of the objection, as the snake-bite example shows. Sometimes no amount of conscientious information-gathering or efforts to discern likely consequences can guarantee an intuitively right act. There is no way to ensure failing to bring about devastation, short of requiring well-motivated agents to have true beliefs about the consequences of their acts.

This in turn suggests an even deeper worry. Slote’s claim that genuinely well-motivated agents of necessity engage in conscientious attempts to discern the consequences of their acts comes close to conceding that rightness depends on more than just motives. Requiring well-motivated agents to be conscientious about calculating likely outcomes seems needed only because there is pressure for agent-basers to ensure convergence between good motivation and our intuitive judgments of rightness—which, in some cases at least, depend on what happens outside the agent’s mind.

This is not to say that the only reason to require admirably motivated agents to be conscientious about discerning likely consequences is because of our intuitive sense that rightness depends on

\textsuperscript{25} Slote 2001: 18.
outcomes. It does seem that being admirably motivated is incompatible with certain forms of
gross negligence in calculating consequences. But avoiding gross negligence is a pretty low bar.
We shouldn’t preclude spontaneous displays of love or compassion from counting as well-
motivated because they don’t involve calculating likely outcomes.

If the main reason to require well-motivated agents to discern likely consequences is to establish
a reliable connection between being well-motivated and bringing about acceptable outcomes,
Slote’s account of good motivation appears tailored to capture the importance of outcomes for
deontic judgments. It is, I think, preferable simply to admit that what goes on outside our minds
is relevant for rightness. This suggests that agent-basing, like virtue-subjunctive and targeted
theories, makes otiose considerations about virtue rather than more evaluatively basic
considerations—the good outcomes at which virtuous motives typically aim—the determinants
of rightness.

One line of response to these difficulties is to tone down the ambitiousness of agent-basing, and
in particular to convert it from a theory of all-things-considered rightness (or wrongness) into a
theory of pro tanto rightness (or wrongness).\(^\text{26}\) Pro tanto agent-basing says that one of the
factors that contributes to making acts right or wrong is the motives from which they are
performed. An act is pro tanto right just in case it’s performed from good motives, and pro tanto
wrong just in case it’s performed from bad motives. This would let agent-basing escape the
counterexamples, by allowing it to affirm that outcomes as well as motives are a determinant of
rightness. If the importance of outcomes can outweigh the importance of motives, well-
motivated acts can be pro tanto right and all-things-considered wrong, and poorly-motivated acts
can be pro tanto wrong and all-things-considered right. And this seems sufficient to
accommodate the counterexamples.

While this strategy meets the objections, it does so by abandoning the element of agent-basing
that distinguishes it as a distinctive alternative to other theories of rightness. Agent-basing loses
its distinctiveness, and with it much of its interest, when it stops making claims about all-things-
considered rightness.

---

\(^{26}\) I’m grateful to Gopal Sreenivasan for this suggestion.
Perhaps a more promising reply from agent-basers involves clarifying the sense of ‘rightness’ they intend their theory to explain. There is, in particular, ambiguity in agent-based theories over whether the sense of ‘rightness’ they have in mind is supposed to be subjective or objective. Indeed I have been unable to find examples of agent-basers who even explicitly recognize this distinction, never mind take a stand on whether they intend their view as an explanation of objective or subjective rightness.

It seems to me that the examples I’ve marshalled pretty clearly show that objective rightness does not depend on good motives. No matter how well-motivated you are, giving me the wrong antidote and leaving trails of destruction in your wake will not, on any plausible view of objective rightness, be right in light of the facts. But now suppose that agent-basers insist that this is perfectly compatible with their theory, since it is subjective rightness, not objective rightness, they intend to explain. Since the acts in the examples (non-culpably giving me the wrong antidote, etc.) are likely to be subjectively right, they are not counterexamples to the idea that subjective rightness depends on being well-motivated.

While this is true, there are other examples that do show that being well-motivated does not suffice for doing the subjectively right thing. These examples can be found in chapter two, where I argued that it is possible for ATC well-motivated agents to do subjectively wrong acts. These same examples can be used to show that subjective rightness does not depend on being well-motivated, contrary to what subjective versions of agent-basing claim.

4.4. \[ P (rightness-based\; virtue) \succ P (virtue-based\; rightness) \]

I’ve been arguing that three prominent virtue-based theories of rightness suffer serious defects. Virtue-subjunctivism struggles to do justice to our sense that the responses of virtuous agents demand an explanation, if we are to regard them as normatively binding. In a similar vein, targeted theories invert the most plausible order of explanation, claiming that complex considerations about whether seemingly basic evaluative facts (relieving suffering, etc.) count as the targets of virtues explain the evaluative significance of those facts, rather than vice-versa. Agent-based theories return the wrong verdicts about particular acts, and fail to capture the intuitive importance of facts outside the agent’s mind for determining deontic status.
These considerations should lower our credence in the truth of virtue-based rightness. Whether they should lower our credence in virtue-based rightness relative to VA depends on whether VA is subject to similar difficulties. But VA is not subject to these difficulties. With respect to the theory of rightness, defenders of VA are just committed to their being some workable notion or other of objective rightness. Defenders of VA can disagree over which acts are objectively right; over what makes objectively right acts right; over exactly which factors are evaluatively significant; over what makes these factors significant; and so on. By acknowledging that certain features of the world are evaluatively significant apart from facts about virtue, VA avoids the basic problems with virtue-based theories of rightness. Given this, the shortcomings with virtue-based theories of rightness pointed out in this chapter give us reason to prefer VA.

Whether or not we should all-things-considered prefer VA to virtue-based theories of rightness depends on whether there are good reasons to reject the claim that an act’s virtue depends on its objective rightness. At this point in the dialectic, however, opponents of rightness-based virtue are running out of options. The most natural way to argue against the idea that the virtue of acts depends on their being objectively right is to try to show that it’s false that acts are virtuous only if they’re objectively right. I have, however, already met the most promising forms of these objections, namely, the P-R view that acting from good motives suffices for virtuous acts; the claim that good motives plus subjective (not objective) rightness suffices for virtuous action; and the consequentialist attempt to explain virtuous action entirely in terms of the good. None of these views, I’ve argued, provides compelling grounds for rejecting the view that acts are virtuous only if they’re objectively right.

Once we accept that acts are virtuous only if they’re objectively right, the most promising way to deny that virtue depends on rightness is to try to show that rightness depends on virtue; to defend a form of virtue-based rightness. If, as I’ve argued in this chapter, virtue-based theories of rightness are implausible, and having accepted that acts are virtuous only if they are objectively right, we seem warranted to accept that virtue depends on objective rightness. After all, if we can conclude that wherever there is a virtuous act there is a right act, and if we can also conclude that the rightness of acts does not depend on their virtue, the most plausible remaining
explanation is that the virtue of acts depends on their rightness. Once we see the implausibility of virtue-based rightness, it’s difficult to see what else could explain the fact that acts are virtuous only if they are right. My defense of VA’s clause (ii) is therefore complete.
Chapter five
Motive and context

5.1. Virtue and ‘admirable’
The first four chapters constitute an extended argument for clause (ii) of VA: the claim that acts are virtuous only if and (partly) because they are all-things-considered objectively permissible. This fourth and final chapter is devoted to clause (i): the claim that acts are virtuous only if and (partly) because they are all-things-considered well-motivated.

I said in the Introduction that I was not going to argue for the claim that acts must be well-motivated to be virtuous. The only person I know who denies this, Julia Driver, does so by appealing to cases in which badly-motivated individuals perform acts that serendipitously realize large amounts of good. It seems to me that such cases support the possibility that acts performed from bad motives may be all-things-considered right. But I don’t take these cases to support the conclusion that badly motivated acts can be virtuous, since I have a strong sense that the acts in question are not virtuous, and that they fail to be virtuous precisely because they are done from bad motives. I thus don’t know how to argue against Driver without begging any questions, and I suspect the same will be true for any view that denies clause (i). For this reason I’ll assume that acts must be well-motivated to be virtuous, and limit my discussion to what counts as good or virtuous motivation.

One thing I will not discuss in this chapter is what a motive is. Throughout I will assume that a motive is a mental state or combination of mental states with both representational and conative elements. This is meant to cover a range of plausible candidates, including beliefs, desires, combinations of beliefs and desires (the ‘Humean theory of motivation’), belief-desire hybrid

---

1 I have in mind Driver’s example of the Mutors (Driver 2001: 55). I am more sympathetic to Driver’s claim that virtue does not require knowledge or true belief about at least some of the features of acts, including their moral status. If successful, I take her arguments on this point to show, not that acts can be virtuous without being well-motivated, but that being well-motivated does not require the true beliefs in question. The claims in this chapter are, I think, compatible with a range of views about the role of true beliefs in virtuous motivation. See Driver 2001 and also Arpaly 2003.
states (‘besires’), and so on. I do not want my assessment of good motives to preclude substantive views in moral psychology about what a motive is.

I will, however, assume that it is only occurrent motivational states that comprise our overall motives for any particular act. It is only occurrent states on which the ATC evaluative status of motives supervenes. Merely dispositional states and the absence of certain states, I will assume, are not part of the grounds for determining ATC good motivation.

I will, more substantively, address two questions in this chapter. The first is what counts as an individual good motive: What are the necessary and sufficient conditions on some inner state being a good motive? The second is what counts as all things considered good motivation: What are the necessary and sufficient conditions for some act to be ATC well-motivated?

These questions are related but ultimately distinct. On the one hand, it is difficult to see how someone’s motivation in performing some act could be ATC good without containing at least one individual good motive. But it does not follow from the fact that someone’s overall motivation contains an individual good motive that it is ATC good. Indeed in some cases this seems false.

Suppose that Sam is moved by an individual good motive—say, the motive of sympathy—to help the shady-looking men struggling to get the statue down the museum steps at 3am. It does not follow that Sam’s act of helping is ATC well-motivated. If Sam should have known that the two men are thieves—or if Sam knows this but just doesn’t care—his overall motives in helping are unlikely to be ATC good, even if his motive of sympathy is good.

I will approach the question of individual good motives and the question of ATC good motivation via different routes. Perhaps the two most influential accounts of ATC good or virtuous motivation in the West have been Aristotle’s and Kant’s. According to Aristotle, virtuous motivation consists in performing virtuous acts ‘for themselves.’ According to Kant,

2 Aristotle also claimed that virtuous motivation consists in desiring acts *dia to kalon*—for the fine or noble. Below I distinguish a moralized and non-moralized interpretation of Aristotle’s
virtuous motivation consists in acting from the motive of duty. I broach the topic of ATC good motivation by engaging these time-honored views below. With respect to the goodness of individual motives, I proceed via the more Russian view that there are multiple forms of admirable motivation, corresponding to multiple right-making features of acts. I will argue for two claims: (i) that the evaluative status of individual motives as good or bad varies with context, and (ii) that no version of the Kantian and Aristotelian slogans identifies a necessary condition on ATC virtuous motivation, though they may state sufficient conditions.

I’ll start with a Russian framework for identifying individual good motives, but first we need a bit of backdrop on virtue-ascriptions and motives. As I said in the Introduction, I take virtue-ascriptions to imply claims about admirability. To claim that someone’s motives are virtuous is to commit one’s self to the judgement that those motives are good, in a way that justifies admiration. I will use an intuitive, commonsense notion of ‘admiration’ and take considered judgments about the admirableness of motives to be dependable guides to determining the virtue of motives.

Given this framework, one natural question is how admirable motives must be to count as virtuous. Up till now I’ve been able to ignore questions about how good something must be to warrant virtue-ascriptions. This question didn’t arise in my defense of clause (ii) because permissibility is a threshold notion. It makes no sense to ask how permissible acts must be to count as virtuous; they must, on VA, simply be permissible. The question has application here because goodness of motives, unlike permissibility, comes in degrees.

It seems to me that any and all admirable motives are so far forth virtuous, to whatever extent that they are admirable. Thus, a slightly admirable motive will be slightly virtuous, a moderately admirable motive will be moderately virtuous, and so on. This lowers VA’s bar for virtuous action, permitting acts that are only very slightly admirably motivated to count as virtuous.

---

view, to which the to kalon formulation and the ‘for themselves’ interpretation (respectively) correspond.
Just *how virtuous* these acts will be is another question. Defenders of VA have two options here. The first is to make the degree of virtue of virtuous acts simply a function of the degree to which their motives are good. The second is to make the degree of virtue of virtuous acts a more complex function of the degree to which their motives are good and facts about the nature of the permissible success (e.g., the goodness of the end realized). Either of these views is compatible with my claim that any and all admirable motives are so far forth virtuous, in the sense relevant to clause (i) of VA. For now that’s enough; as I said in the Introduction (footnote 6), I’m going to remain neutral on the best way for VA to explain the degree of virtue of virtuous acts.

5.2. *A Rossian framework for individual good motives*

One helpful framework for thinking about individual good motives is Rossian in flavor. The Rossian framework, as I’ll understand it, is characterized by four distinctive features.

The first is *pluralism* about right-makers: there are multiple features of acts that count in their favor and contribute to making them right, none of which is reducible to the others. This is unlike utilitarianism, for example, which insists that the rightness of acts depends just on their utility.

The second feature is right-maker or *reason-invariantism*: if some feature of an act (such as that it will keep a promise or benefit another) contributes to making an act right in one context, that feature *always* contributes to making acts right, whenever it is instantiated.³

The third feature is what I will call reason-motive *correlationism*: there is an equivalence relation between individual reasons for action and individual good motives, such that some feature of an act counts as a reason for action if and only if it is good to be moved by it.

---

³ By a ‘right’ act I’ll just mean one that’s sufficiently supported by reasons, whether or not these reasons are ‘moral’ in some special sense (I do not know how to make out a clear distinction between ‘moral’ and other reasons, or whether we should try). In this sense of ‘right’ it’s trivial that right-makers give us reasons for action. With this stipulation I gloss over one of the most enduring questions of Western philosophy: the question of why we should be moral. Those who want to use a more loaded sense of ‘right’ may want to read my claims about reasons in what follows as claims about explicitly moral reasons.
Conjoined with right-maker invariantism, correlationism entails the fourth feature of the Rossian framework, *motive-invariantism*: if an individual motive is good (neutral, bad) in one context, it is good (neutral, bad) everywhere. The evaluative valence of motives does not vary with context.

This latter claim will occupy the lion’s share of my focus. I’ll argue that it is false that an individual motive that is good in one situation is good in every situation. Whether an individual motive is good is contextually variable, in a way that does not let us infer from the fact that it is good in one context that the same motive is good in other contexts. If I am right about this, an interesting disjunction follows, which I discuss at the end of the chapter. Either particularism about reasons is true, or there is not, as many prominent philosophers have assumed, a tight connection between having a reason for action and the goodness of being moved by it.⁴ First we need to consider the Rossian framework.

According to Ross, we are obligated to perform several types of acts, where the duty to perform acts of one type is not reducible to the duty to perform acts of any other type. Ross puts this in terms of *prima facie* or *pro tanto* duties.⁵ While it would take us too far afield to unpack the precise meaning of this phrase, we can at least say that an act’s being a *pro tanto* duty counts in favor of the act’s being all-things-considered right. The basic Rossian claim is that we have several *pro tanto* duties, and that we must often—perhaps always—weigh and balance these against each other when determining what to do.

Suppose we have a correct list of *pro tanto* duties, as follows:

*Pro tanto duties*

- Keep promises
- Increase the good
- Make reparation for wrongs

---

⁴ See e.g. Korsgaard 1986; McDowell 1995; Smith 1994.
⁵ Ross himself speaks of *prima facie* duties, but it seems widely admitted that he had something closer to *pro tanto* duties in mind.
Etc.

That an act is the keeping of a promise, or increases the good, or makes reparation for wrongs, counts in favour of it, and tends to make it right. Given that these are our duties, the generic justifying features of actions will be:

*Generic justifying features of acts*
- That the act keeps a promise
- That the act increases the good
- That the act makes reparation for wrongs
- Etc.

The generic justifying features of acts are the properties on which *pro tanto* rightness depends—the *pro tanto* right-making features of acts. To call these features ‘generic’ is to say that they invariably contribute to making acts right, counting in favor of acts wherever they show up.

On the view in question to be well-motivated is to desire to perform acts because, or from a belief that, they have the generic right-making or justifying features. For every generic justifying feature of acts, there is a different type of good motive corresponding to it. This yields yet a third list, this time of good motives:

*Good motives*
- Desiring to perform an act because it keeps a promise
- Desiring to perform an act because it will increase the good
- Desiring to perform an act because it makes reparation for wrongs
- Etc.

To have a good or virtuous individual motive is, on this view, to have a motive on the correct *Good motives* list, which itself consists in being motivationally sensitive to a *pro tanto* right-making feature of acts.
One way to be sensitive to a right-making consideration is to recognize it as a reason or right-maker, under an explicitly moral description. But this is not, I will assume, necessary for having a good motive. If the fact that an act will make Sam happy speaks in favor of that act, making it pro tanto right, it is pro tanto good to be moved by a desire to make Steve happy, even without the further thought that making Steve happy is a reason or right-making feature. To have a good motive is to desire an act because it has a property that makes it pro tanto right, whether or not one takes that property to make it pro tanto right. I will also assume that being moved by a right-making feature involves having a belief (or belief-like state) that the act in question has that feature, but not necessarily that this belief is true. The desire to perform an act from a belief that it will prevent suffering or keep a promise can be a good motive, even if it is in fact false that the act in question will prevent suffering or keep a promise. This said, unless otherwise noted, I’ll restrict the discussion that follows to cases where the relevant beliefs are true.

Because the considerations that count in favor of acts are, on this view, generic or invariable, it follows that the motives on the Good motives list are likewise invariable; motives on the Good motives list are good whenever they are instantiated. We find this in Ross himself. According to Ross, while the value of a good motive varies with its intensity, the moral status of motives—their goodness or badness—remains fixed. This can be seen in his claim that a person who is moved to perform some act because it will benefit one person, but who is at the same time indifferent to the pain that act will cause others, nonetheless acts from a good motive. Ross says that in this case the person’s motive to benefit the one is purely good, though the action is not. As with right-makers or pro tanto duties, an individual motive that is sometimes good is, for Ross, always good. Such invariably good individual motives may not suffice for ATC good

---

6 I intend ‘belief-like state’ to cover a range of plausible candidates for the kinds of states involved in moral judgments. These will be any states capable of having the right sort of representational content, and might include perceptions, belief-desire hybrid states, and even (on some views) desires themselves. That is, I don’t want the framework to preclude substantive views in moral psychology.

7 It is worth noting that nothing I say should be taken to indicate that motives are the only inner states that can be virtuous or vicious. As Ross himself recognized, certain desires, pleasures and pains, and reactive attitudes that are not motives can be virtuous or vicious.

8 Ross 1930 (ch. 7) and 1939 (ch. 12).

9 Ross 1939: 306-308.
motivation, but they always contribute positively to our overall ATC motivation, making it pro tanto better.

5.3. Motive and context
My critique of the Rossian framework takes as its target motive-invariantism, the idea that the evaluative valence of motives is fixed. If acting on a particular motive—say, the motive to keep a promise, or repair a wrong—is good in one context, is it good in all others? Suppose, as seems plausible, that the motive to keep a promise is sometimes good. If the motive to keep a promise is good in some contexts, motive-invariantism entails that it is always good. But now consider the following type of case.

Conflicted
You have promised a friend that you will pick her up from the airport. On your way you see a car in the ditch, mangled, someone’s arm hanging limply from the driver’s side window. No one else is around. You see that if you stop to help, you’ll be late, perhaps very late, to meet your friend. You are motivated to keep driving—this will, after all, let you keep your promise and pick up your friend on time. But you have a stronger motive to stop and help—help is, after all, urgently needed, and on-balance it’s the right thing to do. Your motive to help wins out, and you pull over to do so.

Compare this case with a second.

Wholehearted
The situation is the same as before, only this time you are not motivated to keep driving, even if it will let you keep your promise. You non-conflictedly and wholeheartedly pull over to help.

Is the motive to keep a promise good in Conflicted?\(^{10}\)

\(^{10}\) Certain Kantians might deny that the motive to keep a promise is, in Conflicted, really a motive at all, since it is only those inclinations that we endorse and on which we act that, strictly speaking, count as our motives. See Herman 1981. I cannot here adjudicate different approaches to what counts as a motive. I assume a commonsense approach, which I think will
If motive-invariantism is true, and the motive to keep your word is, as I’ll assume, good in some contexts, the motive that tempts you to keep driving is good. The motive to keep a promise may not suffice to make your motives overall or all-things-considered good, but it does contribute positively to the overall goodness of your motives, making them *pro tanto* good. It is so far forth good to feel some attraction toward continuing past the wreck, and so some conflict toward helping, by recognizing that your word is at stake. This, however, has a seemingly false implication. To take a step back, if the *individual* motive to keep a promise is always good, someone with this motive should have better overall motivation, other things equal, than someone who lacks it. Of course, other things may not be equal. But when they are—when the only difference between cases is the presence or absence of the motive in question—our overall motives are better when they contain an invariably good motive than when they do not. Presumably this is part of what it means to say that a motive is invariably good. Since *Wholehearted* differs from *Conflicted* only by lacking a (putatively) good motive, other things are supposed to be equal in these two cases. It would then follow that the motives in *Conflicted* are better than the motives in *Wholehearted*.

The problem is that the person in *Wholehearted* seems no less well-motivated for wholeheartedly helping. This is meant to be modest. I am not saying that the motives in *Wholehearted* are *better* than the motives in *Conflicted*; just that they are not worse. I find this hard to deny. Someone who wholeheartedly helps people in need, in a situation where it is clearly right to do so, has no reason to regret the fact that she was not conflicted, or to try to improve her motivational performance by adding desires for clearly wrong alternatives the next time she is in a similar situation. 11 It just does not seem true that we improve our motivational structure in cases like this by adding conflicting motives. 12

---

11 Though it may be appropriate to regret other features of the situation; see below.

12 Which is not to say that we can *never* improve our motivational performance in this way; see below.
To re-emphasize, saying that the motive to keep a promise is not good in this context does not imply that it is positively bad. It can fail to be good by being neutral. Neither does it imply that someone who conflictedly helps is not all-things-considered well-motivated. All-things-considered good motivation may be compatible with significant inner conflict. What is at issue is rather whether the individual motive to keep a promise is good in this context. My claim is that insisting that it does conflicts with our sense that in cases of resolvable conflict it is sometimes no worse wholeheartedly to do what is right than to do so conflictedly.

It’s worth noting that this sense is not limited to cases of conflict between the specific values in the cases above. Conflicted and Wholehearted involve a resolvable conflict that pits keeping a promise against helping people in urgent need, but the variantist intuition does not depend on it being these values, rather than others, that conflict. Suppose that instead of keeping a promise you are on your way to make someone happy; or to make reparation for a wrong; or to perform an important service for the local ecosystem. Assuming that it is sometimes good to be motivated by these things, it will, on motive-invariantism, be good to be moved by any of them to continue past the wreck. This seems false. It doesn’t seem better to be conflicted by any of these motives than to be conflicted by the motive to keep a promise. We have variantist intuitions not just about cases where urgent need conflicts with keeping a promise, but where urgent need conflicts with many weaker reasons or duties.

We also have variantist intuitions about cases of conflict with no urgent need.

Conflicted*

Bill and Ted are equally moderately happy. You must choose between preventing something that will result in extreme suffering for Bill, and making Ted happier. Imagine that there is no urgent need in this case; you have a good window of time in which to make the choice. You are moved to make Ted happier—it would be nice to see Ted even happier. But you feel a stronger motive to prevent Bill’s suffering—he would greatly suffer, and it is clearly the right thing to do. Your motive to prevent suffering wins out; you conflictedly prevent Bill’s suffering.
Wholehearted*

Same as Conflicted*, only you wholeheartedly prevent Bill’s suffering.

Insofar as making another happy is sometimes a good motive, motive-invariantism implies that your motives in Conflicted* are better than your motives in Wholehearted*. This again seems false. Having the conflicting motive, while perhaps not making your overall motives worse, does not seem to make them better.

While urgent need is missing from this example, it shares a common feature with the first: both are cases of resolvable conflict where it is clear what should be done. This might tempt us to conclude that wholeheartedly doing what’s right is just as good as conflictedly doing so whenever it’s clear what should be done in cases of resolvable conflict. If that is true, motive-invariantism will fail to capture a lot of cases. Of course, it might not be true. We might not always have a strong sense that wholeheartedly doing what’s right is just as good as conflictedly doing so when reasons conflict but it’s clear what should be done. But it is likely that we will often have this sense. All such cases will be marks against motive-invariantism.

We are also likely to have variantist intuitions about scenarios involving no conflict at all.

Choking Friend Plus

Your friend is choking on the far side of the lawn. You see that she is in danger and start to sprint toward her. As you run across the yard you realize that saving your friend will, in addition to letting you save her life, also let you repair a minor wrong. (You have recently wronged her by breaking a promise to help her move across town.) In response you experience an additional desire to help—this would, you figure, be an effective way to patch things up.

Choking Friend Minus

Same as Choking Friend Plus, minus the motive to repair the wrong by saving your friend.
If the motive to repair wrongs is sometimes good, it is on motive-invariantism better for you to save your friend partly out of a desire to repair a wrong, rather than solely because her life is at stake. In that case (and assuming everything else is held equal) your motives in *Choking Friend Plus* are better than your motives in *Choking Friend Minus*. This seems false. Helping your friend solely because her life is at stake does not seem to betray a motivational shortcoming. Indeed the desire to repair a wrong seems to introduce an element of pettiness into your overall motives, one that is missing in someone motivated solely by the good of her friend. While this pettiness might not suffice to disqualify your motives as all-things-considered good, it does support the conclusion that your motives in *Choking Friend Minus* are at least as good as your motives in *Choking Friend Plus*.  

There are two general ways for motive-invariantists to resist the challenge posed by these cases. The first is simply to insist that the overall motives that contain the putatively good individual motive (the overall motives in *Conflicted*) are better than the overall motives that lack it (the overall motives in *Wholehearted*). Unsurprisingly, this looks to me like bullet-biting. A more sophisticated rebutting strategy agrees with the argument to this point that the overall motives in both cases are on a par, but explains this in a way that is compatible with the relevant individual motive being good, as motive-invariantism holds. The strongest version of this approach will both give positive reasons for thinking that the individual motive remains good and advance a compelling explanation for why the overall motives in *Wholehearted* are just as good as the overall motives in *Conflicted*, even though the latter contain a good motive the former lack. This is the rebutting strategy I’ll consider in what follows.

---

13 This is reminiscent of Bernard Williams’ famous case involving ‘one thought too many’ (though in Williams’ example, unlike here, the offensiveness of the extra thought is supposed to have something to do with its explicitly moral content). Williams 1981: 18.
14 Someone might be tempted to object to this example on the grounds that it is doubtful whether saving your friend from choking really will let you repair the wrong in question. I share these doubts. But I deny that it undermines the example, since motives based on false beliefs can still be good. Even if I am wrong that baking you this cheesecake will make you happy, my desire to do so may still be admirable. Moreover, to say that the motive to repair a wrong (etc.) counts as good when it is based on true beliefs about an act’s ability to repair wrongs (etc.), and not otherwise, is at least to flirt with, and may even require, the sort of motive-variantist assumptions for which I am arguing.
5.3.1. *Comparisons?*

One straightforward reason for thinking that the motive to keep a promise must be good in *Conflicted* is that we need this assumption to make sense of other cases. Consider a situation where the motive that tempts you away from helping is worse than the motive to keep your word. If there is nothing good in *Conflicted* about the motive to keep a promise, how can someone tempted by a worse motive have overall worse motives?15

The objection requires us to specify a motive that is worse than the motive to keep a promise. To make the challenge as strong as possible, we should not assume that what tempts you to drive by the wreck is an intuitively bad motive, for example, the desire to ridicule the person you’ve promised to meet. While it’s indisputable that this motive is worse than the motive to keep a promise, motive-variantists aren’t committed otherwise. The fact that the motive to keep a promise is not good in this context does not entail that there are not worse motives in the situation. Motive-variantists can simply say that the motive to ridicule the person you are meeting is worse than the motive to keep a promise because the motive to keep a promise is in this context neutral, while the motive to ridicule is bad.

What if the objection uses cases where the motive is supposed to be, not bad, but neutral? Suppose that what moves you to continue past the wreck is the thought that you will be able to have dinner with your friend at your favorite restaurant once you get her from the airport. (If you do not think this is a neutral motive, feel free to replace it.) Now suppose the objector insists that this motive is worse than the motive to keep a promise. If there is nothing good about the motive to keep a promise, how can being tempted by it be better than being tempted by the motive to have dinner? The objection claims that variantism lacks the resources to deliver this conclusion.

Again, however, it’s not clear why variantists would accept that the motive to have dinner is, in this context, worse than the motive to keep a promise. The most plausible motive-variantist analysis of *Conflicted* claims that the motive to keep a promise is neutral. Since neither of two neutral motives can be worse than the other, the objector must think that the motive to have dinner

15 I’m grateful to Tom Hurka for this objection.
dinner is positively bad. But variantists need not agree. Considered in its own right, the motive to have dinner is not bad, but neutral. Indeed this is what the objection began by assuming! Motive-variantists need not accept that either of the two motives is worse than the other.

A second response concedes that the motive to have dinner is, in this context, worse than the motive to keep a promise, but points out that it doesn’t follow that the latter is good. The motive to keep a promise can be better than the motive to have dinner in two other ways. (1) In this context the motive to keep a promise is neutral, while the motive to have dinner is bad. (2) In this context the motive to keep a promise is bad, while the motive to have dinner is even worse. Something like (1) in particular seems plausible on motive-variantist premises. To conclude otherwise is to make an invariantist assumption: that the status of the alternate motives remains fixed. Motive-variantists can reasonably maintain that, in the cases at issue, both the motive to keep a promise and the motive to enjoy dinner switch down a notch: the former to neutral, the latter to bad. If we share the sense that there’s nothing good about being moved to drive by the wreck, and that there’s something worse about the motive to have dinner, this is just what we would expect.

Either of these options is a viable way to turn back the objection. These strategies also seem capable of generalizing, allowing motive-variantists to capture comparative judgments about motives in a wide range of cases, without invariantist assumptions. Indeed reflecting on this objection helps us to appreciate the resources available to motive-variantism for capturing judgments about particular cases. Once we recognize these resources it seems implausible that motive-invariantists will be able to defend their view by appeal to cases alone. What motive-invariantists need is a theoretical basis for judgments about individual good motives, one that can compellingly ground their treatment of the disputed examples.

Toward this end, motive-invariantists can start by pointing out that in cases of resolvable conflict something important is lost. You are, in Conflicted, breaking a promise to a friend. While morally justifiable, this is still undesirable. In this light someone conflicted about helping might look sensitive to the badness of breaking her word in a way that a wholehearted person is not. If it’s appropriate for concern with foregone goods to be reflected in our motives, conflicted
motives will be better in cases of resolvable conflict. They will do greater justice to the goods that are lost. We can explain the goodness of conflict in these cases with the idea that it is better to divide our motivational energy up between competing genuine goods than to focus it all on one good, to the neglect of the other. This is the heart of an attractive and powerful idea about good motivation: the ideal of proportionality.

There are different ways to formulate the ideal of proportionality. According to goodness-based statements, ideal motivation consists in being moved in proportion to the goodness of the objects of our desires (or other intentional attitudes). Rightness-based statements, by contrast, say that ideal motivation consists in being moved in proportion to the weight of our right-makers or reasons for action. It is also possible to hold mixed views, which are attractive to those who deny that all right-makers are grounded in goodness.\(^{16}\) On any of these views ideal motivation consists in being moved in proportion to the strength of independent evaluative properties. Disproportionate responses to values or reasons—being moved by them not enough, or too much—constitute motivational shortfalls, making one’s motives at least less good than they could be, and in some cases positively bad.\(^{17}\)

Proportionality presents a challenge to motive-variantism only when it’s combined with substantive claims about the existence of reasons. This is because proportionality merely claims that if there are reasons for action, it is good to be moved by them, in proportion to their strength. Proportionality says nothing about if or when reasons actually exist. In this way proportionality differs from the idea, expressed in the first correlationist entailment, that the very existence of reasons depends on it being good to be moved in certain ways. Because proportionality states no conditions on the existence of reasons, it is compatible with reason-variantism. On reason-variantism, proportionality entails that it is good to be moved by a feature when it counts as a reason, in proportion to its strength, and not good to be moved by it when it does not. In that

---

\(^{16}\) See Hurka 2001: chapter 7. In what follows I discuss proportionality within a reason-based rather than goodness-based framework. Since the fact that something is good typically gives us reasons to act in certain ways toward it, I don’t think this comes to much.

\(^{17}\) Hurka 2001 defends proportionality (see chapter 3 in particular). Adams 2006 argues against it (see chapter 2). For an assessment of the debate, Bradley 2007.
case certain motives will be good in some contexts but not others. That is, proportionality conjoined with reason-variantism entails motive-variantism. This shows that proportionality can be wielded against motive-variantism only on the reason-invariantist assumption that certain reasons, those by which it would be good to be moved, are present in the relevant cases. Proportionality presents a challenge to the motive-variantist treatment of Conflicted only if you continue to have reason to keep your promise once you encounter the wreck.

The fact that proportionality poses a challenge to motive-variantism only on reason-invariantist assumptions makes it a non-ideal tool in this debate. One of the things we may want to conclude from the arguments for motive-variantism is precisely that reasons are variable. Of course, those who are dug into reason-invariantist positions will be reluctant to do so. Such people, and proponents of the Rossian framework in particular, would I think be shrewd to cast their lot with proportionality, so far as rebutting the motive-variantist challenge. Doing so forces defenders of motive-variantism to evaluate the overall case for proportionality: its strength and weaknesses so far as capturing judgments in the realm of motives; its simplicity and explanatory power; the possibility that we might possess non-inferential justification for the idea that it’s always or necessarily good to be moved in strict proportion to our reasons for action; and so on. This is a sizable project, forcing us to confront (among other things) deep issues in moral epistemology as well as the proper role of theory in ethics. Obviously I cannot settle this here. What follows will, however, constitute at least the start of a case for thinking that appeal to proportionality is not an effective way for reason-invariantists to turn back motive-variantism. This on two grounds.

First, motive-invariantists who appeal to proportionality as a way to ground the idea that the relevant individual motives are good must still find a way to explain why the overall motives of the person in Wholehearted are just as good as the overall motives of the person in Conflicted. If the idea that the motive to keep a promise is good in Conflicted commits us to saying that someone who helps wholeheartedly is overall worse motivated than someone who helps conflictedly, we will, it seems to me, have reason to prefer a motive-variantist rather than proportionality analysis of these cases, given the undesirability of this thought. (Mutatis

---

18 See Hurka 2001 for the most extended treatment.
mutandis for the other cases.) In fact I do not think that motive-invariantists have a compelling way to capture these judgments about overall motives. This is the topic of the next section.

Second, I noted earlier that one motivation for proportionality—indeed perhaps the prime motivation—is the thought that it lets us capture the importance of foregone alternatives in cases of resolvable conflict. On this approach desirable competing alternatives necessarily demand positive motivational responses from us, at the time of our choice. But this is a substantive assumption, one that, while intuitive, is not on reflection obviously true. It may be that the importance of foregone alternatives does not always demand positive motivational responses at the time of choice, but is at least sometimes sufficiently reflected in appropriate post facto reactive attitudes—certain types of regret and restitution, for example—that motive-variantism has no trouble delivering. If the motive-variantist treatment of the cases can do equal justice to foregone goods, an important motivation for proportionality will be lost. I return to this thought below. My reasons against using proportionality to turn back motive-variantism will not be complete until then, when this piece of the puzzle becomes clearer.

5.3.2. Variantism debunked?
I’ve been considering reasons to think that the individual motives in the relevant cases—the motive to keep a promise in Conflicted; the motive to repair a wrong in Choking Friend; etc.—remain good. This is the first part of a two-part rebutting strategy, which begins by providing principled reasons to accept that the relevant individual motives are good and then tries to accommodate our judgment that, nonetheless, the overall motives in the relevant pairs of cases are on a par. If as motive-invariantism claims the relevant individual motives (to keep a promise, repair a wrong, etc.) are good in the relevant contexts, it is, as I claimed earlier, natural to conclude that the overall motives that contain these individually good motives will be better than the overall motives that lack them.

This falls out of the seemingly indisputable idea that part of what it means to call an individual motive good is that, other things equal, its presence makes our overall motives better than they would be without it, in conjunction with the stipulation that everything else is in fact equal in the relevant pairs of cases. In that case the overall motives of someone tempted away from helping
by the motive to keep a promise (for example) will be better than the overall motives of someone who wholeheartedly helps. To block this unpalatable conclusion motive-invariantists must explain why this is not, contrary to appearances, so. Motive-variantism has a simple explanation for why the conflicted and wholehearted motives are on a par: the conflicting motive (to keep a promise) is not good in this context. Motive-invariantists must in essence debunk the variantist explanation, by providing a better one. They must say that, if we scrutinize the cases carefully enough, the stipulation that other things are in fact equal doesn’t hold up: we will be able to find aspects of the cases beyond the presence or absence of the relevant individual motive that can explain why wholehearted motives are on-balance just as good as conflicted motives, even if the latter include a good motive the former lack.\footnote{Again, mutatis mutandis for cases that do not involve conflict.}

One way to do this is to find an intrinsically bad element of motivation that is present in the overall motives that contain the (putatively) good individual motive, but absent in the overall motives that lack it. Focus for now on the original pair of cases, Conflicted and Wholehearted. If we can locate an intrinsically bad element in Conflicted but not Wholehearted, the presence of a good motive will not necessarily make the conflicted motives better, since the goodness of the individual motive may be offset by the intrinsically bad element. Such a strategy is, of course, only as strong as the bad-maker it proposes. What could the intrinsically bad element in Conflicted be? The cases are structured so that the only thing that differs between them is the presence of the motive; we move from Conflicted to Wholehearted merely by subtracting the motive to keep a promise. Given this, the intrinsically bad element in Conflicted must be tied very closely to, and most likely explained by, the motive to keep a promise.

One candidate for such a bad-maker is motivational conflict itself. Conflict is present in Conflicted, because the motive to keep a promise is present; conflict is absent in Wholehearted, because the motive to keep a promise is absent. This is just what we are looking for; conflict itself might explain a difference between the cases. The problem is that conflict itself is a weak candidate for the work that needs to be done. Recall that we are looking for an \textit{intrinsically bad} feature that can offset the value of an intrinsically good motive. But in many cases the bare fact of conflict is not intrinsically bad. An extreme counterexample is the tragic moral dilemma.
There is nothing intrinsically bad about Agamemnon being conflicted over whether to sacrifice Iphigenia or let his ships languish in the harbor. There would be something wrong with him if he wasn’t conflicted! Or suppose someone has to choose between letting several close friends die or letting a major metropolis and all of its inhabitants be destroyed. We would not think well of this person if she did not experience a good deal of conflict over letting her friends die, even if she knows it is the right thing to do.

This response might seem unavailable to those who accept the motive-variantist treatment of the cases. Why are competing motives good in extreme moral dilemmas, but not in *Conflicted*? The sentiment behind this objection seems to be that motivational conflict must be good in one case (*Conflicted*) because it is good in another (Agamemnon). But this reflects a motive-invariantist assumption: that if conflict is sometimes bad, it is always bad. Motive-variantists should, absent further argument, be skeptical. If the valence of particular motives can change, it would not be at all surprising if the valence of motivational conflict changes as well. The question for us should instead be whether conflict is bad in the cases we’re considering, those used to motivate variantism. While it seems plain that the motives in *Wholehearted* are at least as good as the motives in *Conflicted*, it is not at all clear (indeed it seems to me false) that there is something intrinsically bad about the latter. Rather than disparage the motives in *Conflicted*, the motive-variantist interpretation lets us affirm that wholeheartedly helping is just as good as conflictedly doing so without committing us to thinking that there’s something positively bad about being conflicted between helping and keeping your word.

A second proposal for the bad-maker we are looking for is a lack of sufficient intensity in particular motives. Since we are looking for an intrinsically bad feature of motivation that varies with the presence of the (putatively) good individual motive, this lack of intensity must be entailed by the presence of this motive. In *Conflicted*, for example, the strategy must say that someone drawn away from helping by the motive to keep a promise is, by virtue of this drawing away, not as concerned about helping as she should be. In *Choking Friend Plus* we must similarly say that the presence of the motive to repair a wrong entails that the motive to save your friend is insufficiently strong. Either way, since being insufficiently concerned is a
motivational shortcoming, we have a way to derive the conclusion that the overall motives in the cases are on a par, without needing to accept a variantist interpretation of them.

The trick here is to make out a case for thinking that the mere presence of the relevant motive necessarily renders you insufficiently concerned. Consider first cases of resolvable conflict. The only way that having a conflicting motive of necessity leaves you insufficiently concerned is if doing the right thing necessarily demands all of our motivational energy. If that is true, any motive that pulls us away from the right will of necessity leave us inadequately concerned for the right. Motives that draw us away from the right could then be intrinsically good but fail to make our overall motives better. They will make our motives pro tanto better, by being intrinsically good, but pro tanto worse, by siphoning away concern for the right. Similar things apply to Choking Friend. While you are not in this case tempted away from doing what’s right, it might yet be thought that your friend’s well-being requires all of your concern, such that any concern for repairing a wrong of necessity leaves you deficient in this regard.

This strategy requires us to say both that the individual conflicting motive is good and that being conflicted is worse than being wholehearted in the relevant cases. While perhaps not incoherent, this is an uncomfortable combination of ideas. Even if we can make good sense of it, it is suspect for the same reasons that warranted rejection of the idea that conflict itself is bad. It does not seem true that doing what’s right necessarily requires all of our concern, or that being conflicted is always bad. Again, the fact that you are conflicted in very difficult moral dilemmas does not make your motives any less good. Likewise, when there are multiple reasons supporting an act, as in Choking Friend, it seems false that we are of necessity motivationally deficient if one of these reasons does not capture all of our motivational energy. We may be; but this will depend on details of the case. If you fixate on repairing the minor wrong as you sprint to help your friend, such that this begins to crowd out concern for her well-being, something

---

20 It might seem strange to hold that one and the same motive makes our overall motives both pro tanto better and pro tanto worse. While this sounds paradoxical, we can defuse the worry by noting that what makes our overall motives worse in these cases is not the motive per se—that would be incoherent—but rather the lack of sufficient concern it entails. While lack of sufficient concern is, on the view in question, entailed by the presence of the good conflicting motive, the two are not the same.
does seem amiss with your motives. But this crowding out is a contingent fact about the situation. Imagine another case. You have promised to help your friend, and it is the right thing to do. The time to help arrives, and you are motivated to keep your word. On your way you also realize how much your friend will benefit from your help. You acquire a second motive: to benefit your friend. Imagine that the motive to benefit your friend does not make your motive to keep your word any weaker, and that you would have kept your promise even if the motive to benefit were absent. Such a case seems perfectly possible. But of course acquiring the motive to benefit your friend does not make your motives worse.

Since there is little reason to think that acquiring an additional motive entails a corresponding weakening of motive elsewhere, or a weakening below what is required, the current strategy seems doomed to failure. If ideal motivation doesn’t require all of our concern to be reflected in single motives, there will be times when it is no less good to divide our motivational energy up in less single-minded ways. In these cases sufficient concern will not exhaust our available concern, and our leftover concern can find appropriate expression in other motives. If so we cannot appeal to the badness of insufficient concern to offset the goodness of individual motives.

A final debunking strategy runs as follows. Rather than look for an intrinsically bad feature of the overall motives that contain the (putatively) good individual motive to offset its goodness, we look for an intrinsically good feature of the motives that lack the (putatively) good motive to make up for its absence. Instead of looking for a bad feature of Conflicted to offset the goodness of the motive to keep a promise, we look for a good feature of Wholehearted to make up the difference in the other direction.

Keeping in mind that what separates the cases is just the presence of the individual motive, it is difficult to see what this intrinsically good feature might be. Perhaps the best chance for this approach is an analogue of the one just discussed, where facts about the intensity of motives play the role of difference-maker. Rather than there being something intrinsically bad about failing to have single-minded concern in the relevant cases, perhaps there is something intrinsically good

---

21 The goodness of dividing our motivational attention up between competing goods may, contrary to what proportionality assumes, be variable.
about having it. In *Conflicted* this strategy requires us to say both that a conflicting individual motive is good and that it is better to help wholeheartedly than conflictedly. (Mutatis mutandis for *Choking Friend.*) But again, even if we can make sense of this, it runs into the rebuttals just raised. It is not intrinsically good for agents in the thick of difficult moral dilemmas to exhibit single-minded concern toward one or another alternative. Neither is it intrinsically good for you to be single-mindedly fixated on keeping your promise, to the exclusion of the well-being of your friend. This debunking strategy also seems doomed to failure.

While there may be other ways for motive-invariantists to explain our sense that the overall motives in the relevant cases are on a par, it is not easy to see what they are. And even if motive-invariantists can find a way to get this conclusion, it is difficult to imagine that their explanation could be simpler or better than that the relevant individual motives simply fail to be good in these contexts. Indeed we are likely to suspect that invariantists who insist on finding other explanations when there is one as simple as variantism right in front of them are being driven by commitments outside the theory of motives. This is of course fine, though it is important to be clear about the price of these commitments when it comes to evaluating motives. Those of us who want to bracket (for the time being) our other commitments and assess the debate between motive-variantism and motive-invariantism in its own domain, on the basis of how well each view does at capturing considered judgments about particular motives, will I suspect have far less trouble accepting that motive-variantism makes better sense of the cases.

I will return to the Rossian framework and the implications of motive-variantism below. First I want to spend some time on ATC good motivation, which I will approach by way of Aristotle and Kant.

5.4. *Aristotle and Kant*

I’ve been focusing on the goodness of individual motives, arguing that individual motives can be good in some contexts but not in others. As I said in the beginning, the question of what counts as an individual good motive and what counts as ATC good motivation are, while related, distinct. In particular, it does not follow from variantism about individual motives that we cannot state general and non-variable conditions on ATC good motivation. In this section I’ll
approach the goodness of ATC motivation via Aristotle and Kant. While I cannot do justice to the whole range of considerations relevant here, I will sketch some of what I take to be the major issues and challenges to these views. I will argue that we can glean sufficient conditions—but no necessary conditions—for ATC good motivation from the Kantian and Aristotelian slogans.

Aristotle claimed that being virtuously motivated involves performing virtuous acts ‘for themselves,’ while Kant identified virtuous motivation with the motive of duty. I will assume that these are claims about the ATC goodness of motives: Aristotle and Kant are saying that acts are all-things-considered well-motivated just in case they are done from duty (Kant) or ‘for themselves’ (Aristotle).

There are, however, at least two interpretations of these slogans: a moralized interpretation and a non-moralized one. The moralized interpretation identifies desiring an act ‘for itself’ with desiring an act because it is virtuous (under the description ‘virtuous’), and identifies the motive of duty with the desire to do an act because it is right (under the description ‘right’). The non-moralized interpretation identifies desiring an act ‘for itself’ with whatever it is about the act that makes it virtuous, and acting from the motive of duty with desiring acts for whatever it is about them that makes them right.22 Let’s start with the moralized interpretations.

5.4.1. Moralized Aristotle and Kant
The moralized interpretations say that an act is ATC well-motivated just in case it is done from the thought that it is dutiful (in the case of Kant) or virtuous (in the case of Aristotle). It seems to me that neither moralized interpretation states a plausible necessary condition on ATC good motivation. If it did, it would not be possible to be ATC well-motivated without being motivated by thoughts about duty or virtue. This seems false. Desiring to relieve someone’s suffering or bring them joy often seems to suffice for ATC good motivation, even if there is no explicit connection to thoughts of duty or virtue.

It might be objected that someone motivated by relieving suffering or bringing joy would not count as ATC well-motivated unless they also believed that doing so was right or virtuous, or at least not wrong. We must, the objector claims, be sensitive to the moral status of acts in order to count as ATC well-motivated.

This may be true. I doubt, however, that this sensitivity must form part of our occurrent motives for them to be ATC good. It may, if it involves an occurrent belief that an act is virtuous or right which figures as part of the explanation for why we act. But this belief might also be motivationally inert: a background or dispositional belief about the act, without being our reason for doing the act, or figuring in our occurrent motives at all. Those who want to insist that we must be sensitive to the moral status of acts in order to be ATC well-motivated must explain why this sensitivity must form part of our occurrent motives, rather than just functioning as a filter that helps make certain counterfactuals about us true (that we would not perform the act were it wrong, etc.) without being part of what actually moves us.²³

Whether desiring acts because they are right or virtuous suffices for ATC good motivation is, it seems to me, more difficult. The main issue here is whether there are times when performing acts because they are one’s duty or virtuous betrays an unhealthy preoccupation or ‘fetish’ with moral concepts, rather than the particularities of concrete persons and relationships.²⁴

This question has been debated at length. I will rest content to point out that we needn’t think that acting from the motive of duty or virtue is the best motive we could have to think that it suffices for ATC good motivation. It seems clear to me that there are times when acting for the sake of duty or virtue is not the best motive we can have. It seems better to save your partner or child from a fire because you love her than because it is virtuous or your duty. But it is far less clear to me that doing so from the motive of duty or virtue does not suffice for ATC good

²³ Cf. Herman 1981. Herman argues that ‘empiricist’ views of motives that identify motives with desires will not be able to say that certain motives are both present and inert. That may also be true. But note that I am talking here about the belief that some act is virtuous or right or not wrong, not a desire to do the act because it is right or virtuous. It seems to me perfectly possible to believe that an act is right without that belief forming part of one’s occurrent motivation.

²⁴ The criticism has been around at least since Schiller’s famous indictment of Kant. More recently it has found compelling expression in Stocker 1976; Williams 1981; and Smith 1994.
motivation. Intuitions will vary, but I am inclined to think that acting from the motive of duty or virtue will typically suffice for ATC good motivation, at least when we are right or not negligent in believing that the act in question really is right or virtuous.

5.4.2. Non-moralized Aristotle and Kant

The non-moralized interpretations of Aristotle and Kant seem to me more interesting. They claim that desiring acts ‘for themselves’ or from the motive of duty is equivalent to desiring acts for the properties that make them virtuous or right. To be distinct from the moralized interpretation, we should assume that the properties in question, those that make acts virtuous or right, are not moral properties, such as virtue or rightness itself. Rather, the non-moralized interpretations say that motivation is ATC good just in case it is based on a belief that the act in question has the non-moral properties on which its virtue or rightness depends.

Let’s take the non-moralized Aristotelian slogan first. It says that acts are ATC well-motivated iff they are done for the properties that make them virtuous. This can be a difficult claim to evaluate. Many (including Aristotle) hold that what makes acts virtuous is at least partly facts about the motives from which they are done. This implies that virtuous acts must be motivated at least partly by thoughts about their motivation to be ATC well-motivated. The most plausible version of this view will, I think, also hold that acts are made virtuous because they have properties that are not tied to their motivation: because they relieve pain or cause happiness, bring about just outcomes, or whatever it is about them (apart from facts about their motivation) that makes them worth doing. The view would then say that we are ATC well-motivated just in case we are motivated to do acts partly for the properties that make them worth doing (apart from

---

25 See Dancy 1993, Stratton-Lake 2000, and Markovits 2010, each of whom makes hay in this context over whether rightness itself gives us reason to perform acts. Of course, this question or something like it has been a perennial worry throughout the history of moral philosophy, from Thrasymachus to Hobbes to Prichard to early Foot.

26 Whiting 2002. On views where the virtue of acts depends just on their motivation, the non-moralized Aristotelian view would say that acts are virtuously motivated iff they are motivated just by facts about the agent’s motives (since it is solely facts about motives that, on these views, make acts virtuous). But this can’t be the whole story. We still need to know what the relevant first-order motives are—those motives which it will be virtuous to endorse and be moved by. The non-moralized Aristotelian view doesn’t give us this information. So it is a bad fit for views on which virtuous action depends just on motivation; or at least it requires supplement.
facts about their motivation) and partly because we are or can be motivated in this way, by the good-making properties of acts.

It is the second conjunct that presents difficulties. To say that we must be motivated by our motives seems to mean that we must take some sort of higher-level pro-attitude toward our first-order motives: we must endorse the content of our first-order motives, or perhaps endorse the fact that we have them (or both), and take them as good reasons for acting as we do. We must not only want to do acts from certain first-order desires—for whatever it is (apart from facts about motivation) that makes them worth doing—but also do acts at least partly because we can do them, or are doing them, from these first-order desires.

This is a strange form of motivation. For one, it is likely to be quite rare: I doubt many people are motivated by thoughts about their first-order desires. More importantly, this simply does not seem necessary to be ATC well-motivated. Someone motivated to help a friend solely out of concern for her well-being can be ATC well-motivated, even if she is not moved by thoughts about the goodness of her first-order concern.

If we assume that acts are made virtuous at least partly by facts about motives, then, the non-moralized Aristotelian slogan does not seem to state a necessary condition on ATC good motives. Might it provide a sufficient condition? I think it will depend on how we feel about being moved by our own desires. Such motivation can seem excessively self-occupied. Someone who is moved by the goodness of her own concern for a friend, rather than just by the well-being of her friend, can seem unattractively self-absorbed.²⁷

But here again I think it’s important to make a distinction between the best motivation and ATC good motivation. If we equate virtuous motivation with the best motivation, the non-moralized Aristotelian view will likely strike many as failing to provide a sufficient condition on virtuous motives, because of the self-absorption involved in being moved by our own first-order desires. But since VA denies that virtuous motivation must be the best, defenders of VA can recognize

---
²⁷ Cf. Hurka 2013, who argues that excessive preoccupation with self is a foundational feature and failure of Aristotelian virtue ethics.
that being moved by one’s first-order desires is not the best form of motivation without needing to say that it disqualifies motives as ATC good. While the self-absorption involved in being motivated by the thought that we can act from certain first-order desires seems to me to detract from the overall goodness of motives, this does not seem to me so egregiously self-absorbed as to disqualify motives as ATC good.

So far I have argued that the moralized interpretations of Aristotle and Kant do not provide necessary condition on ATC good motivation, but that they may very well provide sufficient conditions; and that the non-moralized interpretation of Aristotle does not provide a necessary condition on ATC good motivation. Whether the non-moralized Aristotelian slogan states a plausible sufficient condition is more difficult. In the rest of the section I will, however, focus on the non-moralized interpretation of the Kantian slogan. My take on whether the non-moralized Aristotelian version states a sufficient condition will emerge below.

The non-moralized Kantian slogan says that acts are ATC well-motivated just in case they are done for the properties that make them right. As it stands, this statement of the non-moralized Kantian view is too ambiguous to assess. The ambiguity is over whether we should interpret it as requiring ATC good motivation to be based on the features of acts that in fact make them right, or the features of acts that we believe make them right. While the former interpretation is possible, I assume that most proponents of the view do not intend it, since it results in an extremely strict view of ATC good motivation. If ATC good motives requires us to desire acts for features that in fact make them right, acts must (obviously enough) actually be right in order to be ATC well-motivated. But this seems pretty clearly false. It seems possible to exhibit good motives when performing acts that aren’t right, since sometimes even the most admirably motivated acts turn out, through non-culpable ignorance or plain bad luck, wrong.

---

28 This is closely tied to the distinction between objective and subjective rightness. But I think it’s simpler and more to the point to frame things as I do in the text. Interestingly, this distinction is rarely made by those who discussed the non-moralized Kantian view.
The second interpretation is more promising. It says that acts are ATC well-motivated just in case they are done from beliefs about right-makers: just in case we are motivated by the features of acts that we take to justify them and make them right. For example, suppose that Melody has been bitten by a poisonous snake and needs the antidote. You believe the following: (i) I have strong reason to help Melody; (ii) the way to help Melody is to give her the antidote; (iii) the syringe contains the antidote. Suppose you take these three beliefs to justify giving Melody the syringe. The view under consideration would then say that you are ATC well-motivated just in case you are moved to do so on the basis of believing (i)-(iii). This is so even if giving Melody the syringe turns out to harm her rather than help her, and is wrong in light of the facts. I’ll focus on this interpretation of the view in what follows.

Is it true that we are ATC well-motivated just in case we are motivated by beliefs about an act’s right-makers? One difficulty here involves saying just which of our beliefs suffice to justify acts and make them (from our perspective) right. Talk of a right-making relation seems to denote a sufficiency relation broadly akin to logical entailment. This suggests that the relevant beliefs are all of those needed to justify the act and make it right from our perspective.

The first worry is that there are likely to be several beliefs involved in this relation, and that it is implausible to require ATC good motives to be based on all of them. For example, in the case of Melody above you will at least typically take more than the fact that Melody needs help, and that you can help her by giving her the syringe, to justify the act. We can see this by adding uncertainty over whether other considerations outweigh your reasons to help.

Suppose you suspect that Melody is a crazed sociopath with a functional nuclear warhead in a nearby cave, who intends to fire it that afternoon on a nearby city. You do not outright believe that this is true of Melody; but neither do you believe that it is not. Rather, you are epistemically split down the middle. Under these conditions you wouldn’t think that giving Melody the antidote is justified. The fact that you think it just as likely as not that Melody will recover and employ the warhead makes it reasonable to doubt that you should give her the antidote, to the point where you might justifiably refuse doing so.
What this reveals, it seems to me, is that you never took giving Melody the antidote to be right \textit{just} on the basis of your beliefs that Melody needs help and that the way to help her is to give her the syringe. You \textit{also} assumed that there are no countervailing considerations that would make saving Melody wrong. When pushed to provide a complete justification for why giving Melody the antidote is right, at some point you would inevitably list your belief that there is nothing of comparable importance that would be lost by doing so. Moreover, something similar seems true of every act we take to be justified. When pushed, what we take to justify any particular act is not just that we believe it has certain positive right-making properties that count in favor of doing it (that it would save Melody) but also that it \textit{fails to have} properties sufficient for wrongness (that it would \textit{not} result in catastrophe).

The problem for the non-moralized Kantian view is that it does not \textit{always} seem necessary to be moved by all of these beliefs—both our beliefs about positive right-makers and our beliefs about the lack of wrong-makers—in order to be ATC well-motivated.\footnote{\textit{Mutatis mutandis} for the non-moralized Aristotelian view, at least insofar as multiple features of acts can be involved in making them virtuous.} Suppose that you desire to make your partner happy, and believe that bringing her flowers will do so. On this basis you desire to buy her flowers. This desire need not, it seems to me, be based on the belief that buying the flowers has no wrong-making properties in order to amount to ATC good motivation. You need not pause at the florist to consider the possible ways in which giving flowers might turn out wrong in order for your motives to count as ATC good.\footnote{This shouldn’t be taken to show that these acts \textit{have no} wrong-making features. The fact that buying your partner flowers doesn’t have catastrophic consequences is, plausibly, a necessary condition on its being right. It’s just that being admirably motivated doesn’t require you to act on recognition of this fact.} We need not always be actively looking over our shoulders, or moved by beliefs about lack of wrong-makers, to be ATC well-motivated.\footnote{Again, this is compatible with the belief that there are no wrong-makers being present but motivationally inert in the well-motivated person.}

It’s worth noting that the basic difficulty here is likely to afflict the view even if beliefs about wrong-makers are excluded. Even if we find a principled way to deny that beliefs about the absence of wrong-makers are part of what we take to justify acts, the view in question seems too
demanding. We can see this by considering cases where the act in question is justified by beliefs about multiple positive right-making features.

Suppose that you must choose between spending Saturday helping your son, and spending Saturday helping your daughter. You believe (correctly) that by helping your son you can benefit him slightly more than you can benefit your daughter by helping her. But suppose it is also true that you have recently wronged your daughter, breaking a promise to take her fishing. You believe (again, correctly) that spending Saturday helping her would be an appropriate way for you to repair this wrong. Suppose you conclude on this basis that you should help your daughter rather than your son.

Note here that your belief that you can repair the wrong by helping your daughter may not suffice to explain why you should help her. If you believed that helping your daughter would not benefit her at all, or would benefit her far less than helping your son, helping your son would (let’s suppose) take precedence over helping your daughter, since you can help him more. Here you take helping your daughter rather than your son to be justified partly because of your belief that you can benefit your daughter only slightly less than your son, and partly because of your belief that doing so will (in addition) let you repair the wrong. If this is right, the non-moralized Kantian view requires you to help your daughter from both of these thoughts in order for you to be ATC well-motivated.

The difficulty is that it seems overly strict to require you to help your daughter for both of these reasons in order to exhibit ATC good motives. Helping her either on the basis of your desire to benefit her, or on the basis of your desire to repair a wrong, may be enough. This is (again) not to deny that it would be better to help her for both right-making properties. If you believe that spending Saturday with your daughter is right both because it will help her and because you have wronged her, but are only moved to help her by (say) a desire to repair the wrong, that is likely to indicate a lack of sufficient concern with helping her—which, plausibly, detracts from the goodness of your motives. Nonetheless, being moved to spend the day with her in order to repair a wrong may still be ATC good. It seems to me that we should not set the bar for ATC good motives so high as to rule out cases like this. (Indeed we could press this argument further by
coming up with cases where an act is justified on the basis of five or six different sorts of considerations, but where you are only moved by four or five of them. But I take it the point is clear enough.) Some acts that we take to be justified by more than one right-making feature may be ATC well-motivated without every right-making feature figuring as grounds of our motives.

So it seems to me that we should reject that ATC good motivation requires being moved by all of the features of acts that we take to justify them and make them right. That is, the non-moralized Kantian view does not provide a necessary condition on ATC good motives. Might it supply a sufficient condition? Does being moved in this way always amount to ATC good motivation?

The main challenge to this idea is the possibility of mistaken moral agents: people who perform acts that are in fact wrong but that they believe are right. Some such cases involve people motivated by repugnant moral beliefs. Take someone, call him Tim, who believes that the interests of women or blacks or Arabs deserve far less consideration than the interests of white North American males. Tim will face situations in which he has to decide how to relate to women or blacks or Arabs. In these situations Tim might believe that he should treat individuals in ways that are in fact disrespectful and wrong, precisely because these individuals do not deserve his respect.

Suppose Tim believes that he should leave the power tools he borrows from his next-door neighbor, Meg, out in the rain. Tim believes this because he thinks that women should not be so bold as to use power tools, and that any woman who uses them needs to be taken down a notch. Since Tim’s leaving Meg’s tools in the rain is motivated by his desire to take Meg down a notch, and since it is this belief—that Meg needs to be taken down a notch—that Tim takes to justify his act and make it right, Tim can, if the non-moralized Kantian slogan is correct, be ATC well-motivated when he leaves Meg’s tools in the rain for this reason. But this looks like bad or vicious, not ATC good, motivation.

Some are likely to insist that we need to hear more about Tim’s mistaken moral beliefs before we conclude that he isn’t well-motivated. Suppose in particular that Tim’s mistaken moral beliefs are epistemically justified. Tim has, we can suppose, grown up in a severely isolated and insular
community. All of his role models have espoused racist beliefs, and have articulated what seem to Tim to amount to a compelling case for the mistaken moral beliefs he endorses. Should we still deny that Tim is ATC well-motivated when he leaves Meg’s tools in the rain because he thinks Meg needs to be taken down a notch, given that Tim does what he believes is right for the reasons that (he thinks) make this act right, and is justified in holding these (repugnant) beliefs? There are three at least three ways to respond. The first is to say that, while acting from a false but justified belief can in principle give rise to good motivation, it is not possible to have justified repugnant moral beliefs of the sort Tim has—in which case Tim is not well-motivated. The second is to accept that Tim’s repugnant moral beliefs may be justified, but nonetheless deny that Tim is well-motivated, on the grounds that being moved by false moral beliefs does not always (or often or ever) amount to having a good motive. The third way to respond is to accept that Tim is well-motivated, since he does the act for the reason he believes makes it right, and this belief of his is justified.

Those who think the non-moralized Kantian view states a sufficient condition on ATC good motives should take the third route. So should those who think the non-moralized Aristotelian slogan states a sufficient condition, since we can run structurally similar cases involving agents who (like Tim) do repugnant acts that they falsely believe are virtuous for properties that they falsely believe make those acts virtuous. But we should accept the third way only if neither the first nor the second option is more plausible. Is the third option most plausible?

The first option, recall, denies that Tim’s repugnant moral beliefs can be justified, and concludes that Tim is not ATC well-motivated. This seems to involve rejecting the general possibility of justified moral ignorance. While it seems to me that there are cases of justified moral ignorance, it also seems difficult to articulate an airtight case for them, and certainly beyond the scope of this chapter. I’ll simply rest content to note that skeptics about justified moral ignorance may have independent reason to reject the claim that Tim is well-motivated, and thus

---

32 So long as some moral ignorance is justified, we will presumably be able to get similar challenges to the view up and running, even if Tim’s own brand of moral ignorance is not or cannot be justified.
independent reason to think that the non-moralized Kantian slogan does not state a sufficient condition on ATC good motivation.

The second approach accepts that Tim’s repugnant moral beliefs may be justified but denies that his motives are ATC good. I think this is more plausible than saying that Tim is ATC well-motivated. It simply does not seem to me true that Tim is well-motivated when he mistreats Meg because he thinks Meg needs to be taken down a notch, even if Tim’s repugnant moral beliefs are justified. That Tim’s repugnant beliefs are justified certainly seems to support the conclusion that we should not blame Tim for holding these beliefs and being motivated in this way. But it hardly follows that Tim is ATC well-motivated, in a way that merits admiration. At best I think we should regard Tim’s case as unfortunate, or perhaps tragic in some deeper sense. If Tim really cannot be reasonably expected to believe much differently, Tim deserves our concern and help and pity, not admiration.

This said, I have no deeper explanation or argument for this sense of mine. And so I have little to say to those who insist that Tim’s motivation is ATC good in this case. I rest content with pointing out that defenders of the view in question are committed to deeming Tim’s sexist motivation, not just blameless or unfortunate or tragic, but downright admirable. Those who see nothing untoward with this implication can continue to hold that the non-moralized Kantian and Aristotelian slogans state sufficient conditions on ATC good motives. Those who, like me, find it incredible to think that Tim merits admiration should conclude that they do not.

By way of recap, I’ve been considering what I take to be the two most prominent accounts of virtuous motivation in the West: Aristotle’s idea that virtuous motivation consists in doing acts ‘for themselves’ and Kant’s idea that virtuous motivation consists in the motive of duty. I started by distinguishing a moralized and non-moralized version of both these views. I then argued that neither the moralized version of Aristotle nor the moralized version of Kant states a necessary condition on ATC good motivation, since some ATC well-motivated acts are not done from explicitly moral thoughts. Nonetheless, I suggested that the moralized version of these views may succeed at stating sufficient conditions on ATC good motivation, even if they do not state plausible sufficient conditions on the best motivation. With respect to the non-moralized
versions I argued that neither the Aristotelian nor Kantian slogans state necessary conditions on ATC good motivation, and that whether we think they state sufficient conditions depends on what we think of motivation based on repugnant (perhaps epistemically justified) moral beliefs.\textsuperscript{33}

It seems to me that the difficulty involved with stating necessary conditions on ATC good motivation coheres with variantism about individual motives. Since the evaluative status of ATC motivation is in large part a function of the evaluative status of the individual motives that comprise it, we should expect a certain amount of variability when evaluating ATC motivation. It would be surprising if variantism about individual motives were true but we could succeed at giving a highly systematized account of ATC good motivation. The fact that the Aristotelian and Kantian slogans struggle and I think ultimately fail to state necessary conditions on ATC good motivation can be taken to reflect variability at the level of individual motives.

At the same time, the fact that the moralized versions of Aristotle and Kant arguably succeed at stating sufficient conditions on ATC good motivation indicates that there are certain dependable forms of good motivation, at least construed at higher levels of abstraction. This should temper the sense of vertigo we might be inclined to experience when considering the variability of individual motives. While being virtuously motivated is complex, it is not so complex as to leave us completely in the dark. Being concerned with the rightness or virtue of acts, and arguably with the properties that make them so (provided we are not horribly morally mistaken), may be reliable guides to being ATC well-motivated, even if the evaluative status of individual motives is variable, and even if there are other forms of motivation that might also be good in these contexts.

5.5. \textit{Implications}

I want now to return to motive-variantism, and to unpack an interesting implication of that view. I began the paper by laying out the Rossian framework, which consists in four claims: (i) pluralism about right-makers, (ii) right-maker or reason-invariantism, (iii) reason-motive correlationism, and (iv) motive-invariantism. Earlier I argued that (iv) is implausible: variantism

\textsuperscript{33} And, in the case of Aristotle, the degree of self-absorption involved with being motivated by our first-order desires.
captures considered judgments about individual motives that invariantism has trouble delivering. Suppose for a moment that these arguments are successful, showing that variantism outperforms invariantism in the theory of motives. As I said above, it would not follow that we should reject motive-invariantism and with it the Rossian framework. For we may have reason to evaluate the debate between motive-variantism and motive-invariantism on the basis of more than how each view performs in the realm of motives. If motive-variantism has what we take to be unpalatable implications in other domains, and in the theory of reasons in particular, we might ultimately opt for motive-invariantism, even if variantism outperforms it on its own turf.

This possibility is particularly salient given the status of correlationism. If we take correlationism as a fixed point, and granted plausible substantive claims about the reasons and/or motives we (sometimes) have, the preceding arguments can be used to derive variantism about reasons. Indeed all we need is the first correlationist entailment, which says that something is a reason for action only if it is or would be good to be moved by it. If that is true, the features of acts by which it’s not good to be moved in Conflicted and the other examples (that an act keeps a promise, repairs a wrong, etc.) do not in these cases give us reasons for action. And if, as seems indisputable, these considerations sometimes give us reasons for action, reason-variantism follows.

There are three ways to respond to attempts to use motive-variantism to derive reason-variantism. The first option holds the first correlationist entailment fixed and accepts the move from motive-variantism to reason-variantism. This will be appealing to particularists and reason-variantists generally, who can simply welcome the arguments for motive-variantism as additional support for their view. The second option again holds the first entailment fixed but turns the modus ponens into a modus tollens, concluding that motive-variantism must be all-things-considered false, despite its performance in the theory of motives, given what it implies about the variability of reasons. Reason-invariantists and proponents of the Rossian framework are likely to find this route attractive. The third option is perhaps the most interesting. It simply declines to take the idea that it’s always good to be moved by our reasons for action for granted. In so doing it grants a measure of autonomy to debates over the variability of reasons and the variability of motives.
Of these three options the second seems to me the least desirable, at least on the assumption that the preceding cases really are best explained on motive-variantist assumptions. If I am right about this, someone who continues to accept motive-invariantism would at the least be committed to a range of seemingly unacceptable judgments about good motivation.

Things are even worse if, because of invariantism’s inability to return these judgments, variantism about motives turns out to be a better overall approach to evaluating motives. The debate between motive-variantism and motive-invariantism is, after all, about motives. Presumably the aim of a theory of motives is to make sense of the realm of motives: to yield and explain attractive judgments about motives in particular cases. Someone who came to believe that motive-variantism is overall superior to motive-invariantism in the realm of motives, while continuing to accept motive-invariantism on other grounds, would in effect be giving up this aim. This seems tantamount to accepting that our commitments in other areas (and in the theory of reasons in particular) ought to trump our attempts to understand the realm of motives. But in that case we need to know why making sense of the realm of motives is less important than these other commitments. There must be principled grounds for thinking that our other commitments trump the theory of motives. I find it difficult to see what these might be. But perhaps there is a case to be made.

Someone who accepts motive-variantism should, by contrast, take the first route or the third, either accepting variantism about reasons or rejecting the correlationist approach to the relation between reasons and motives. This seems to put those with reason-invariantist commitments in a bit of a bind. Indeed many might think that the third route is no option at all, given the plausibility of the first correlationist entailment (i.e., something is a reason for action only if it is good to be moved by it). Alternately, reason-invariantists swayed by the arguments for motive-variantism might begin to wonder whether we should really take the first correlationist entailment for granted. Indeed on reflection we might find something suspicious, or at least

---

34 Michael Smith calls this a platitude (1994: 150). Mark Schroeder and Stephen Finlay likewise claim that views on which we have reasons for action only if it would be good to be moved by them may be “too weak to be interesting” (2012). Kieran Setiya incredulously asks: “Who would deny that, when the fact that p is a reason for A to φ and A is incapable of being moved to φ by the belief that p, she falls short of rational excellence?” (2012: 7).
surprising, about reason-variantism falling out of motive-variantism so effortlessly. This suspicion is (if nothing else) bolstered by sociological considerations: none of the principal debates over variantism about reasons have, to date, been conducted in the theory of motives.

To end I want to begin to make a case for the claim that the first correlationist entailment should not be granted axiomatic status. For those who accept motive-variantism, one obvious way to do this is to muster arguments for the superiority of reason-invariantism in the realm of reasons, which can combine with motive-variantism to undermine the first correlationist entailment. Of course, there are already many such arguments, and no hope of deciding the debate over reason-in/variantism here. Instead I want simply to sketch some neglected consideration that weigh in favor of reason-invariantism (not necessarily decisively) and that are part of the case for motive-variantism itself.

I said earlier that one motivation for proportionality is the thought that it allows us to capture the importance of foregone alternatives, and that we can remove this motivation by showing how to do justice to these alternatives without appealing to motivational conflict, in a way that’s consistent with a motive-variantist analysis. If motive-variantism can do justice to foregone goods, and if what is needed to capture their importance is none other than reason-invariantist assumptions, motive-variantism and reason-invariantism mutually support each other, and provide at least the beginnings of a principled way to reject the idea that it is always good to be moved by our reasons for action.35

5.5.1. Remainders and reasons

Because some may be prone to reject out of hand the idea that the motive-variantist analysis of the cases could fit well with reason-invariantist assumptions, given what this would imply about the relation between reasons and motives, it may help to begin by noting that the first

---

35 Since my basic strategy will be to argue that in certain cases of conflict it is not good to be moved by a consideration that might nonetheless function as a reason for action, the arguments that follow will not challenge the second entailment. The second entailment says that, if it is good to be motivated by some feature of action, that feature is a reason for action. To conclude on the basis of my strategy that we should reject the second entailment amounts to denying the antecedent.
entailment—the idea that something counts as a reason for action only if it is good to be moved by it—is a substantive claim. We can see this by making a simple conceptual distinction between reasons for action and reasons for motives. Someone tempted to deny that it is always good to be moved by our reasons for action should still hold that it is always good to be moved by our reasons for motives. This latter claim seems undeniable. But since motives and actions are not the same, it is not obviously necessary that reasons for action always give rise to, or do double duty as, reasons for motives. While many may find this idea attractive, it does not seem incoherent to deny it.

Of course, we need grounds to deny it. The line of thought I want to pursue focuses on the idea that the best way to capture the importance of foregone alternatives in cases of resolvable conflict is not in the appropriateness of conflicting motives, as proportionality assumes, but rather in the appropriateness of certain post facto attitudes toward the goods that are lost. What are the appropriate attitudes toward goods that are lost in cases of resolvable conflict?

It’s commonly thought that agents who find themselves in resolvable dilemmas will have a duty to make restitution to anyone who is wronged or suffers as a result of their choice. This thought is especially plausible in cases of broken promises.36 It’s relatively uncontroversial that promise-breakers have duties to compensate the recipients of promises (‘promisees’) for any material losses suffered as a result, where this is taken to include the promisee’s time, energy, money, and so on. The promisee may waive the right to compensation, and we may also want to recognize excusing conditions. Nonetheless, promise-breakers will often be on the hook for material losses.

What about cases where there is no material loss? We can conceive of cases where, as a result of the promise-breaking, promisees benefit: better use is made of their time, energy, money, and so on. In these cases a certain form of restitution still seems demanded. We at least owe the promisee an explanation for why we have broken our word, and a sincere apology for doing so.37

---

36 See in particular Thomson 1990: 91-96.
37 While this perhaps stretches the ordinary sense of ‘compensation’ (it is less clear it stretches the ordinary sense of ‘restitution’), it seems appropriate when we see that the relevant
These claims hold even if we assume that breaking the promise is the right thing to do, as it is in *Conflicted*. The fact that you ought all-things-considered to stop and help the people in the wreck, rather than keep your promise, does not by itself let you off the hook, so far as restitution. If there are material losses, you may still be required to ameliorate them.

Imagine that, because you stop to help, you are late to pick up your friend, and that because of this your friend is late to pick up her dog from the dog-sitter, who makes her pay an extra $100. You should be willing to bear at least some of this cost, given that it is a direct result of your broken promise. There may, of course, be exempting conditions. If your friend is a millionaire who, after hearing the details of your explanation, insists on absorbing the costs herself, you are, plausibly, off the hook. But if your friend is struggling to make ends meet, or simply does not offer to absorb the costs, you should help her shoulder the burden you have caused, even if you do not act wrongly in burdening her. If on the other hand there are no material losses, you must still explain and apologize. Someone who does not explain and apologize for breaking a promise almost always acts wrongly, even if breaking the promise is justified.

A similar set of claims hold for regret, which many take to be a second appropriate response in cases of conflict. As with restitution, different situations call for different kinds of regret. When we resolve conflict correctly, it will obviously not be appropriate to regret having done the wrong thing—to feel remorse or contrition—since we have done what is right. Nonetheless, other forms of regret may be called for. Often in cases of resolvable conflict right choices have bad consequences, as when stopping to help makes your friend late for her dog. In these cases it may be appropriate for you to regret the bad outcomes you cause, even if it is not appropriate to regret the choice all-things-considered, or wish to undo it. This is often called (misleadingly) ‘agent-regret.’ What about when there are no bad outcomes in the ordinary sense?

---

explanation and apology are best thought of as illocutionary speech-acts. There is nothing strange about the thought that you owe someone the performance of some action (repairing the fence, etc.) as compensation for a wrong (vandalizing the fence, etc.).

38 Cf. in particular Foot 1983 and Baron 1988.

39 Williams 1976. It is misleading because the bad consequences that you cause and regret need not be intended by you. In such cases many will think that what you regret is not part of your agency at all.
Imagine, for example, that because you stop to help you are late to meet your friend, but that your friend has fortuitously bumped into an old college roommate at the airport. Their friendship is rekindled and she has a fantastic time. While there are no bad consequences in this case, a measure of regret may still be appropriate. You have, after all, broken your word. Anyone who denies that the normative force of keeping promises can be reduced to its consequences will think that there is something undesirable about breaking promises, even if the consequences aren’t bad. In that case your broken word may be a candidate for appropriate regret.

There are different ways to specify the object of this regret. Again, it will not be regret that you have made the wrong choice, since you have made the right choice. Neither will it be appropriate to regret the choice all-things considered, in a way that entails a wish to undo the choice, since you should not have this wish. Some may nonetheless regard your choice as the proper object of regret. Ross, for example, says that in such cases you may appropriately feel a measure of ‘compunction’ over your choice, since it leaves a pro tanto duty (to keep your word) unmet. Since ‘compunction’ plausibly involves a (mild) form of regret, you should on this view feel some regret over the choice. This strategy depends on it being false that any regret over one’s choice entails a wish to have chosen differently. Those skeptical about this may want to specify the proper object of regret in cases of conflict as the choice-situation, rather than the choice itself: what you may appropriately regret is that the world put you in this bind, forcing you to choose between helping and keeping a promise. Either way many will accept that there are forms of appropriate regret in cases of resolvable conflict, even when we do what’s right and our choice has no bad results.

The appropriateness of restitution and regret gives us a way to recognize the importance of what is lost in cases of resolvable conflict. You should only feel regret over, and make restitution for, the loss of real goods. Because restitution and regret allow us to recognize the importance of

---

40 Ross contrasts compunction with shame and repentance (1930: 28). Shame and repentance are what we should feel when we fail to do our all-things-considered duty, whereas compunction is what we should feel when doing our all-things-considered duty requires us to act against a pro tanto duty. Ross grounds the appropriateness of compunction in the persistence of pro tanto duties (cf. Hurka and Shubert 2012); this is my basic strategy below.

41 Foot 1983.
competing alternatives, motivational conflict is not the only way to do so. If this is right, those who want to use proportionality to cast doubt on motive-variantism must say that, while restitution and regret may allow us to do partial justice to foregone goods, they do not suffice. In addition, motivational conflict is needed. But then we want to know why this is so. Why is the importance of foregone goods never sufficiently captured by restitution and regret? Restitution and regret may, we can recognize, not always suffice. In very difficult or (if there are any) irresolvable moral dilemmas, doing sufficient justice to the alternatives may, as I claimed earlier, require a measure of motivational conflict. But in those cases where wholeheartedly taking one option is just as good as conflictedly doing so, restitution and regret mitigate the need for motivational conflict in the virtuous agent. The current combination of strategies allows us to say that the ideally virtuous agent both wholeheartedly stops to help while making restitution and experiencing appropriate regret over her broken promise. This is everything we really want about the case.

I’ve been assuming that restitution and regret are, in the relevant range of cases, appropriate. What explains this? Where reasons or oughts conflict, but where one ought wins out, many hold that there persists a ‘moral trace’ or ‘residue’ or ‘remainder’ that grounds the appropriateness of restitution and regret. Moral remainders are invoked precisely to capture our sense that morality is not always a zero sum-game; that something important is lost in cases of conflict, even if the right choice is made.

If moral remainders best explain the appropriateness of restitution and regret in cases of resolvable conflict, we have support for a reason-invariantist analysis of these cases. It is difficult to see how a moral residue can persist through cases of conflict unless we have reason to do the acts we forego.

Consider again Conflicted. Suppose that the motive-variantist analysis of it is correct: it is just as good for you wholeheartedly to help the people in the wreck as to do so conflictedly, since the

---

motive to keep a promise is not here good. Someone who (on the basis of the first correlationist entailment) takes motive-variantism to support reason-variantism will conclude that you have no reason to keep your word. But if you do not have reason to keep your promise, where does the moral remainder come from? How can the significance of keeping your word result in a remainder if you have no reason to do it?

It is not easy to see what the answer might be. If it is appropriate for you, rather than anyone else, to make restitution and experience regret, the appropriateness must be anchored in something other than the mere badness of the situation. Your promise must continue to exert normative pressure on you. If the reason-giving force of the promise is canceled, it is difficult to see where that pressure is coming from.

It seems to me that any plausible explanation of the appropriateness of these attitudes, not just appeal to moral remainders per se, will raise roughly the same difficulties. The reason-variantist strategy prohibits us from explaining the appropriateness of restitution and regret in any way that entails that we have reasons to do the acts for which restitution and regret are appropriate. But then explaining the appropriateness of restitution and regret looks to be an uphill task. Those who accept that you have reason to keep your promise can, on the other hand, simply say that the remainder is explained by this fact. Moral remainders seem far easier to explain on the assumption that we have reason to do the foregone acts than on the assumption that we do not.

If persisting reasons for action are required to make sense of restitution and regret, and if the cases used to support motive-variantism are compelling, there will be times when it is not good to be moved by considerations (that an act keeps a promise, makes another happy, etc.) that nonetheless function as reasons for action, grounding the appropriateness of restitution and regret. This would show that the first correlationist entailment is false.

Of course, I have not claimed that the reason-invariantist analysis of restitution and regret is in fact required, or that we have no choice but to accept it. Those who want to accept both motive-variantism and the first correlationist entailment can no doubt look for responses to the preceding arguments. They might, for example, simply deny that restitution and regret are appropriate in
the relevant cases. This would imply that restitution and regret are not appropriate in many cases where they seem to be appropriate, which is unattractive. But maybe there is a case to be made.

On the other hand, those committed to the idea that it is always good to be moved by our reasons for action might get creative with attempts to explain the relevant normative pressure in ways that do not entail reasons for action. I do not doubt that we can find alternate explanations. What seems questionable is that they will be simpler or better than explaining the persisting normative pressure in ways that entail persisting reasons for action. But, again, perhaps there is a case to be made.

Perhaps the most promising line of response recalls the more determinate versions of the first correlationist entailment, mentioned earlier. These versions specify a particular way in which it is good to be moved by our reasons for action, saying that it is always rational or virtuous to be so moved, rather than merely good in some way or other. Defenders of these more determinate statements might find some wiggle room. In particular, those committed to the rationality of being moved by our reasons for action, rather than its virtuousness per se, might insist that what we learn from the preceding arguments (if anything) is just that it’s not always virtuous to be moved by our reasons for action. The person in Conflicted may be no less virtuous for failing to respond to her reason to keep her promise, they might say, but all the less rational for that. I will end with some brief thoughts on this approach.43

The rationalist rebutting strategy faces a two-pronged challenge. First, it must show that the particular motive-evaluations implied by it are plausible. Assuming that the rationalist accepts that in cases of conflict we have reason to do each of two competing acts, they must say that the person who wholeheartedly helps (for example) is less rational than someone who conflictedly does so. This judgment, considered on its own, does not seem just obvious.

43 The converse strategy takes the preceding arguments to show merely that it is not always rational to be moved by our reasons for action, not to show that it is not always virtuous to be so moved. This approach seems less promising to me, simply because I have a strong considered sense that someone who wholeheartedly helps in Conflicted (for example) does not thereby betray a shortcoming in virtue. While I do not have a strong sense that wholeheartedly helping betrays a shortfall in rationality, I do not have a strong sense that it does not, either.
It will presumably be very tempting for the current strategy to ground these judgments in the rationalist version of the first correlationist entailment: the reason it is less rational to help wholeheartedly than to help conflictedly is just that you have a reason to keep your promise, and so a reason to be conflicted. This brings us to the second part of the challenge.

Merely asserting that, while not always virtuous, it is always and necessarily rational to be moved by our reasons for action stifles debate by begging questions against two groups of people: those who simply see nothing whatsoever bad (neither unvirtuous nor irrational) about failing to be moved in the relevant ways, and those who hold that the requirements of rationality can be absorbed into (or just are) the requirements of one-or-another virtue, such as phronesis or prudence. What we want are reasons to accept the rationalist version of the first correlationist entailment. Why exactly should we think that it is always rational to be moved by our reasons for action?

Again, this is different from requesting an explanation for why it is always rational to be moved by our reasons for motives. This demand for explanation does seem misplaced. But to insist that the demand for explanation is misplaced where what is at stake is the rationality of being moved by our reasons for action could not itself be variable. A number of possible views on this score jump readily to mind. Perhaps it is only rational to be moved by certain types of reasons for action, such as those that favor our own self-interest, or the well-being of our nearest and dearest. Or maybe it is only rational to be moved by reasons for action when they are ‘moral reasons,’ or support what is in fact the right or virtuous act. Perhaps on the other hand differences in context, rather than differences in the natures of different types of reasons, make it sometimes rational to respond to reasons for action, sometimes not. The point is not that any of these proposals are likely to be true. It’s rather that their easy availability should deter us from taking the rationalist version of the first correlationist entailment as an axiomatic assumption that can be taken for granted and assumed without argument in this debate.

If this is true, it may turn out to be a welcome development. Faced with the plausibility of motive-variantism, those who are unwilling to budge on the first correlationist entailment may be forced to revisit the grounds for an oft taken for granted metaethical doctrine, which may in turn produce new insights. On the other hand, those who find the preceding arguments for motive-variantism compelling but are willing to hold the status of correlationism open may come to see new avenues in old debates, and in the debate over reason-in/variantism in particular. I have sketched the beginnings of one such avenue here. Either way, it seems to me, debates over the nature of reasons and the nature of motives, and their relation to each other, stand some chance of being advanced. In that case we have an example of how developments in the theory of virtue can impact wider debates in metaethics and the theory of reasons.

5.6. Conclusion

VA claims that acts are virtuous only if and partly because they are ATC well-motivated. In this chapter I have tried to reach some conclusions about what good motivation might look like. I’ve argued that variantism about individual motives is true, which in turn makes it difficult to provide strict necessary conditions on ATC good motivation. Nonetheless, certain versions of the Aristotelian and Kantian slogans may, I suggest, state plausible sufficient conditions. The general lesson is that evaluating motives is a complex and subtle business, though not perhaps so complex as to leave us completely in the dark regarding the content of virtuous motivation. As a general rule, attention to the moral status of acts as virtuous or right, and (under most situations) attention to the non-moral features of acts on which their rightness or virtue depend, will be a reliable guide to being ATC well-motivated.

By way of more general conclusion, I’ve argued in this work that virtuous action is a complex but simultaneously unified notion, combining good motivation and permissible success into an organic whole whose value outstrips the value of its parts in isolation. While it has not formed part of the preceding discussion, I want to end by pointing out that the view I have defended constitutes an extended attempt to show how two basic evaluative properties that are often thought to be radically different and indeed (by some) at loggerheads—rightness and goodness—
can complement each other by forming a single important evaluative entity, which I have argued is none other than virtuous action.

My hope is that this sort of project can encourage us to think more deeply, not just about the nature of virtuous action, but about the nature of goodness and rightness, and the relation between them. If rightness and goodness can be conceived as forming a single evaluatively important entity, this might beckon modest hope for progress in ethics. Many seemingly intractable debates in ethics seem to pit one basic moral concept against another. The debate between deontology and utilitarianism, for example, can be interpreted as a disagreement over whether we can understand the right in terms of the good without doing violence to the nature of the right. But if rightness and goodness can form a unified evaluatively significant entity, virtuous action, while retaining their distinctive characteristics—in a way that allows us to preserve the distinctive features of both the right and the good—why not think that the right and the good, or other seemingly disparate moral properties and concepts, can complement each other in other domains and ways?

I hope that my project here can be taken to exemplify and encourage a broadly conciliatory rather than antagonistic approach to moral theorizing, one that lends itself to more holistic attempts to understand the nature of evaluative properties and the relations between them.
Works cited


Aristotle. *Nicomachean Ethics*.


*Philosophical Studies* 93: 161-188.


Foot, Philippa 1978. “Virtues and Vices.” In her *Virtues and Vices, and Other Essays in Moral Philosophy*. Oxford: Oxford UP.


Hume, David 1739. *A Treatise of Human Nature*.


------------------ 1993. “Putting Rationality in its Place.” In Quinn, Morality and Action, Cambridge UP.


Scanlon 1998. What We Owe To Each Other. Cambridge, MA: Harvard UP.


